First publ. in: Structural Models of Thinking and Learning / (Hrsg.) Hans Spada, Wilhelm Kempf. Göttingen: Hogrefe, 1977, pp. 295-318

A Dynamic Test Model and its Use in the Microevaluation of Instructional Material

Wilhelm F. Kempf

1. Introduction

A traditional goal of curriculum evaluation is the measurement of whether and to what extent curriculum material is producing changes in student behavior according to stated objectives. Although more recently, alternative and broader evaluation strategies have been suggested, achievement testing still plays a central role in the evaluation process.

Despite the fact that it is widely believed that tests can also serve to reinforce student learning, the use of tests in school is often restricted. It is not uncommon to find a dislike for testing among teachers. The present study will suggest, therefore, that instructional material designed for training purposes could simultaneously also be used for the purpose of achievement measurement and that testing could thus be integrated into school teaching as part of regular instruction.

As it turns out, however, joint use of the same material for training and testing calls for a new type of item analysis model. Most of the current test models deal with static situations, assuming that preceding responses of a given person have no influence upon later responses and that the responses of a given person are unaffected by the responses of any other person. If the responses stem from training, however, we have to take learning into account, and the probability distribution of the responses must thus be treated under a dynamic point of view allowing for serial dependence of responses.

In statistical analysis of the data we shall thus have to consider learning effects as well as individual differences in students' achievement and differences in item difficulty. In doing so, we will be interested in finding general empirical characteristics of the training material and to evaluate the material by means of the learning effects elicited and the difficulties inherent in the material.

We thus face the problem of how to make "structural" statements about general empirical characteristics and, at the same time, take individual differences into account. A solution to the problem was provided by RASCH (1966) in his theory of "specific objectivity". It is paralleled in the conditional maximum likelihood method (CML-method) for estimating "structural parameters" of probabilistic models in presence of "incidental parameters" (NEYMAN & SCOTT 1948, ANDERSEN 1973). The following discussion is based on this literature, especially on a dynamic extension of the RASCH model (1960, 1966), suggested by myself (KEMPF 1974a, 1976a).

2. The Model

Let us consider some training material consisting of k items. We define a set of random variables a_{vi} ($v=1,n;i=1,k$) such that $a_{vi}=1$ if student No. v gives a correct response to item No. i and $a_{vi}=0$ if the response is an incorrect one. The response vector $(a_{vi}) = (a_{v1},\dots,a_{vk})$ gives a complete description of the student's responses during training. If $i=1,2,\dots k$ is the sequence in which the items are responded to, the distribution of the response vector follows from Eq. 1

$$(1)\quad p\{(a_{vi})\} = \prod_{i=1}^{k} p\{a_{vi}|a_{v1},\dots,a_{v,i-1}\}.$$

Now let s_{vi} be the partial response vector $(a_{v1},\dots,a_{v,i-1})$. Then we introduce the notion of a conditional item characteristic function

(2) $f_{i.s_{vi}}(\xi_v) = p\{a_{vi} = 1|(a_{v1},\ldots,a_{v,i-1}) = s_{vi}\}$

so that

(3) $p\{a_{vi}|s_{vi}\} = f_{i.s_{vi}}(\xi_v)^{a_{vi}}(1-f_{i.s_{vi}}(\xi_v))^{1-a_{vi}}$,

where ξ_v denotes a latent trait parameter, e.g. the student's (initial) ability in training.

On the basis of this, KEMPF (1974a,b) has suggested a dynamic test model in which the conditional item characteristic functions are assumed to depend on parameters of three types: an individual ability parameter ξ_v, an item difficulty parameter σ_i, and a transfer parameter ψ_s describing how a student's responses are affected by his responses to preceding items. The conditional item characteristic functions of the dynamic model all have the same structural form

(4) $f_{i.s_{vi}}(\xi_v) = \dfrac{\xi_v + \psi_{s_{vi}}}{\xi_v + \sigma_i}$ with $\psi_s < \sigma_i$ for all s and i.

Figure 1 shows the conditional item characteristic in Eq. 4 as a function of ξ and σ for fixed values of ψ. It tends towards 0 when item difficulty σ_i grows to infinity and towards 1 when the student's ability increases towards infinity or when $\sigma_i \to \psi_s$. If $\xi_v \to 0$, the conditional item characteristic function approaches ψ_s/σ_i. Regardless of a student's initial ability ξ_v, after a response sequence $(a_{v1},\ldots,a_{vi-1}) = s$, the student's probability of success on item i will not be less than ψ_s/σ_i.

A full understanding of the role that the transfer parameters play in the model can only be achieved by analyzing how the partial response vectors and the transfer parameters are related to one other.

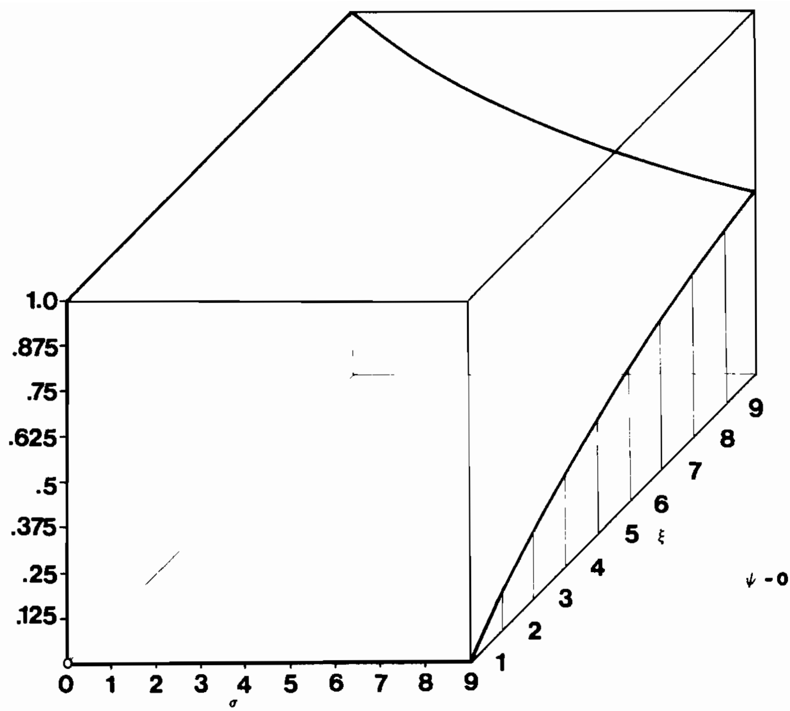

Figure No. 1.: The conditional item characteristic (Eq. 4) as a function of the latent ability variable ξ and of the item difficulty σ at two different values of ψ.

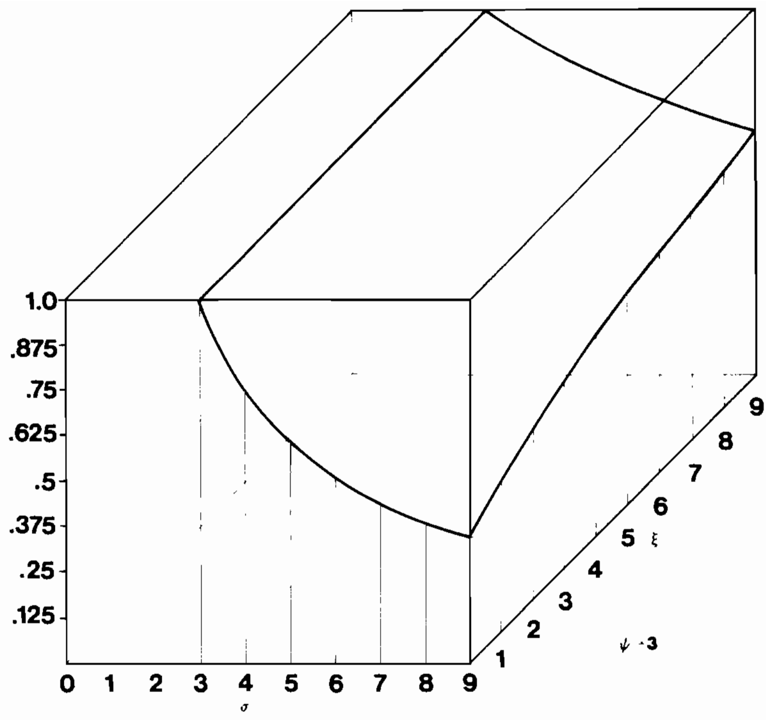

Figure No. 1. continued

Let us assume that we are dealing with some kind of computer assisted training in which the students get immediate feedback about the correctness of their responses. It will be admissable to consider positive feedback as a reinforcement of correct answers, and conditional item characteristic functions can be treated as depending on the number of correct responses to the preceding items

$$(5) \quad r_{vi} = \begin{cases} 0 & \text{for } i = 1 \\ \sum_{j=1}^{i-1} a_{vj} & \text{for } i = 2,3,\ldots,k. \end{cases}$$

According to this assumption

$$(6) \quad f_{i.s_{vi}}(\xi_v) = f_{i.r_{vi}}(\xi_v)$$

holds for all partial response vectors s_{vi} which are compatible with the partial score r_{vi}. The model is reduced to

$$(7) \quad f_{i.r_{vi}}(\xi_v) = \frac{\xi_v + \psi_{r_{vi}}}{\xi_v + \sigma_i} \quad \text{where } \psi_r < \sigma_i.$$

Figure 2 shows this conditional item characteristic as a linearly increasing function of the transfer parameter ψ_r for fixed values of ξ_v and σ_i. It tends towards 1 when $\psi_r \to \sigma_i$ and towards $\xi_v/(\xi_v + \sigma_i)$ when $\psi_r \to 0$. The transfer is a learning effect when ψ_r is an increasing function of r (positive transfer); it is a reactive inhibition when ψ_r is a decreasing function of r (negative transfer); and it is a fluctuation that can be explained by concurrent positive and negative transfers, if ψ_r is a non-monotone function of r.

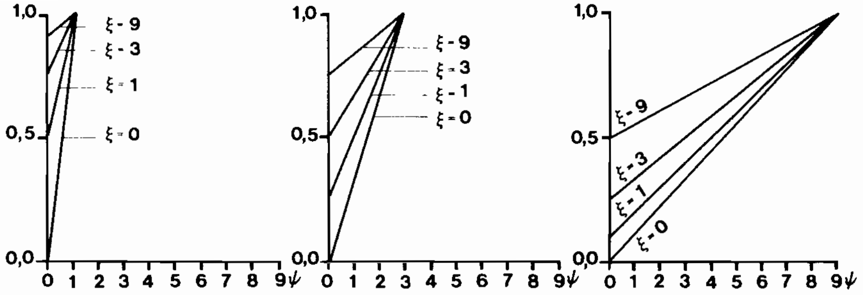

Figure No. 2.: The conditional item characteristic (Eq. 4) as a function of the latent ability variable ξ and of the transfer parameter ψ for fixed item difficulty $\sigma=3$.

However, with n individual parameters $(\xi_1,\ldots,\xi_n)$, k item parameters $(\sigma_1,\ldots,\sigma_k)$ and k transfer parameters $(\psi_0,\ldots,\psi_{k-1})$, the model (Eq. 7) is slightly overparametrized. Note that $(\xi_v + c) + (\psi_r - c) = \xi_v + \psi_r$ and $(\xi_v + c) + (\sigma_i - c) = \xi_v + \sigma_i$. Note further that Eq. 7 remains unchanged if both numerator and denominator are multiplied by some positive constant. Thus, the parameters are measured on interval scales and we can set

$$(8) \qquad \sum_{i=1}^{k} \sigma_i = 1 \quad \text{and MIN } (\psi_r) = 0,$$

which gives a proper parametrization of the model.

If there is no transfer at all, we may set $\psi_0 = \psi_1 = \ldots = \psi_{k-1} = 0$ and Eq. 7 is reduced to

$$(9) \qquad f_{i.r_{vi}}(\xi_v) = \frac{\xi_v}{\xi_v + \sigma_i} = f_i(\xi_v) \text{ for } r_{vi} = 0,1,\ldots,k-1$$

which is the (unconditional) item characteristic function of the Rasch model for binary items.

As KEMPF (1974a, 1976) has shown, the extended model (Eq. 7) has essentially the same mathematical properties as the Rasch model (Eq. 9). The raw scores $a_{vo} = \Sigma_{i=1}^{k} a_{vi}$ are sufficient statistics for the latent ability parameters ξ_v and the item and transfer parameters can be estimated by use of the conditional maximum-likelihood method from the conditional likelihood function

$$(10) \qquad L = p\{((a_{vi})) \mid (a_{vo})\} = \prod_{v=1}^{N} p\{(a_{vi}) \mid a_{vo}\},$$

where $((a_{vi}))$ denotes the matrix of the responses of N students to k items and (a_{vo}) denotes the vector of raw scores of the students. If $a_{vo} = 0$ or $a_{vo} = k$, then $p\{(a_{vi}) \mid a_{vo}\} = 1$ and hence does not contribute to the conditional likelihood (Eq. 10). Students with a raw score $a_{vo} = 0$ or $a_{vo} = k$ can therefore be deleted from the sample.

3. Sufficient Statistics

The raw score a_{vo} is a sufficient statistic for estimating the ability parameter ξ_v <u>if</u> the conditional likelihood of the student's response vector

$$(11) \quad p\{(a_{vi}) \mid a_{vo}\} = \frac{p\{(a_{vi})\}}{p\{a_{vo}\}}$$

does not depend on the student's ability parameter ξ_v.

Inserting Eq. 7 into Eq. 6 and Eq. 6 into Eq. 3 we obtain from Eq. 1 the unconditional likelihood of the response vector (a_{vi})

$$(12) \quad p\{(a_{vi})\} = \frac{\prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r)}{\prod_{i=1}^{k} (\xi_v + \sigma_i)} \prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1-a_{vi}} ,$$

The likelihood of the raw score a_{vo} is obtained by summation of the probabilities $p\{(a^*_{vi})\}$ of all possible response vectors (a^*_{vi}) which are compatible with the score so that $\Sigma_{i=1}^{k} a^*_{vi} = a_{vo}$

$$(13) \quad p\{a_{vo}\} = \sum_{(a^*_{vi})|a_{vo}} p\{(a^*_{vi})\}$$

$$= \frac{\prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r)}{\prod_{i=1}^{k} (\xi_v + \sigma_i)} \sum_{(a^*_{vi})|a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1-a^*_{vi}}.$$

From a theorem by KEMPF & HAMPAPA (1975, 1976), the sum on the right hand side of Eq. 13 can be written in the form

$$(14) \quad \sum_{m=0}^{s} \delta_m(k-s)\, \gamma_{s-m}(k)\, (-1)^m \quad \text{with } s = k - a_{vo}$$

where $\delta_m(k-s)$ denotes the sum of all possible products of m transfer parameters from the set $(\psi_o,\ldots,\psi_{k-s})$, with repetitions

$$(15) \quad \delta_m(k-s) = \begin{cases} 1 & \text{for } m=0 \\ \sum_{j_1=0}^{k-s} \sum_{j_2=j_1}^{k-s} \cdots \sum_{j_m=j_{m-1}}^{k-s} \prod_{t=1}^{m} \psi_{j_t} & \text{for } m=1,2,\ldots s. \end{cases}$$

$\gamma_{s-m}(k)$ denotes the elementary symmetric function of order s-m of the item difficulty parameters $\sigma_1,\ldots,\sigma_k$.

The conditional likelihood of the response vector, finally, is obtained from inserting Eqs. 12 and 13 into formula 11

$$(16) \quad p\{(a_{vi})|a_{vo} = k-s\} = \frac{\prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1-a_{vi}}}{\sum_{m=0}^{s} \delta_m(k-s)\ \gamma_{s-m}(k)\ (-1)^m}$$

Eq. 16 is dependent on the item and transfer parameters only and does not depend on the individual parameters ξ_v. Consequently, a_{vo} is a sufficient estimator for ξ_v, and any extra information about which of the items were answered correctly is useless as a source of inference about ξ_v. However, it can be used for inferring the item and transfer parameters independently from the individual parameters.

4. The Estimation Equations

Let us consider the responses of n individuals with $0<a_{vo}<k$. The students' responses can be arranged in an n x k response matrix $((a_{vi}))$, and the students' raw scores can be arranged in an n-dimensional score vector (a_{vo}). The conditional likelihood of the response matrix given the score vector follows from inserting Eq. 16 into

$$(17)\quad p\{((a_{vi}))|(a_{vo})\} = \prod_{v=1}^{n} p\{(a_{vi})|a_{vo}\}.$$

Now let n_{ri} be the number of students who give an incorrect answer to item number i after $r_{vi}= r$ correct responses to the preceding items $j=1,2,...,i-1$. We observe that, for $i=1,2,...,k$ and for $r=0,1,...,i-1$, the expression $(\sigma_i - \psi_r)$ occurs n_{ri} times in the numerator of the conditional likelihood, each time raised to power $1-a_{vi}=1$. In all other places in which it occurs in the numerator of the conditional likelihood, it is raised to the power $1-a_{vi}=0$. Furthermore, let N_{k-s} be the number of students who give a total of s incorrect responses to the k items so that $a_{vo} = k-s$. Then, we see that, for $s=1,2,...,k-1$, expression (14) occurs N_{k-s} times in the denominator of the conditional likelihood. Hence, we can write

$$(18)\quad L = p\{((a_{vi}))|(a_{vo})\} = \frac{\prod_{i=1}^{k}\prod_{r=0}^{i-1}(\sigma_i - \psi_r)^{n_{ri}}}{\prod_{s=1}^{k-1}\left\{\sum_{m=0}^{s}\delta_m(k-s)\gamma_{s-m}(k)(-1)^m\right\}^{N_{k-s}}}.$$

Finally, taking logarithms of Eq. 18 and setting the first order partial derivatives $\partial\ln(L)/\partial\sigma_\alpha$ and $\partial\ln(L)/\partial\psi_\beta$ (for $\alpha = =1,2,...,k$ and $\beta=0,1,...,k-1$) equal to 0, yields the necessary estimation equations which must be solved under the side conditions $\psi_r < \sigma_i$ $(r=0,...,k-1;i=1,...,k)$:

(19) $\partial\ln(L)/\partial\sigma_\alpha =$

$$\sum_{r=0}^{\alpha-1}\frac{n_r}{\sigma_\alpha - \psi_r} - \sum_{s=1}^{k-1} N_{k-s}\frac{\sum_{m=0}^{s-1}\delta_m(k-s)\ \gamma_{s-m-1}^{(\alpha)}(k)(-1)^m}{\sum_{m=0}^{s}\delta_m(k-s)\ \gamma_{s-m}(k)(-1)^m} = 0$$

for $\alpha=1,...,k$, where $\gamma_{s-m-1}^{(\alpha)}(k)$ denotes the elementary symmetric function of order s-m-1 of the parameters $\sigma_1,..., \sigma_{\alpha-1}$,

$\sigma_{\alpha+1},\ldots,\sigma_k$ and

(20) $\partial \ln(L)/\partial\psi_\beta =$

$$\sum_{i=\beta+1}^{k} \frac{n_{\beta i}}{\psi_\beta-\sigma_i} - \sum_{\substack{s=1\\ s\leq k-\beta}}^{k-1} N_{k-s} \frac{\sum\limits_{m=1}^{s} \gamma_{s-m}(k)\left\{\sum\limits_{j=0}^{m-1} \psi_\beta^j \, \delta_{m-1-j}(k-s)\right\}(-1)^m}{\sum\limits_{m=0}^{s} \delta_m(k-s)\gamma_{s-m}(k)\ (-1)^m} = 0$$

for $\beta=0,\ldots,c_{max}$, where c_{max} denotes the largest observed raw score $a_{vo}<k$.

The technical details for solving the equations numerically will not be discussed here. The interested reader is referred to the papers by KEMPF & HAMPAPA (1975, 1976). A Fortran program for numerical computation of the parameter estimates has been published by KEMPF & MACH (1975). It makes use of a gradient method for solving nonlinear equations suggested by FISCHER & FORMANN (1972). The advantage of the gradient method is its computational simplicity. A disadvantage of the method is that it shows rather slow convergence. As the experience with simulated data has shown, still existing inaccuracies in transfer parameter estimates may be compensated by inaccuracies in the item parameter estimates. The logarithmic likelihood may come very near to its maximum when the parameter estimates still show a relatively strong divergence from the "correct" values. For the goodness of fit test discussed in section 5 and for testing hypotheses about the parameters as discussed in sections 6 and 7 this does not do any harm. In directly interpreting the parameter estimates, however, one should be very careful.

5. A Goodness of Fit Test for the Model

An important step in any statistical analysis is the determination of how well the model fits the data. For the present model, this step can be carried out both in form of a graphical plot and in form of a likelihood ratio test based on an approximation to the χ^2-distribution.

The rationale for the goodness of fit test derives from the conditional approach to parameter estimation. Eq. 16 shows that the conditional distribution of a student's responses, given his raw score a_{vo}, is independent of the latent ability parameter ξ_v and depends on the item and transfer parameters only. Therefore, the latter can be estimated from any subgroup of students G_ν by taking the product of Eq. 16 over the students of this subgroup as our conditional likelihood

$$L_\nu = \prod_{v \in G_\nu} p\{(a_{vi})|a_{vo}\}.$$

If $G_1,\dots,G_M$ are M disjoint subgroups of examinees, we define restricted CML estimates of the item and transfer parameters as the solutions to the restricted likelihood equations

(21) $\partial \ln(L_\nu)/\partial\sigma_i = 0$ for $i=1,\dots,k$,

and

(22) $\partial \ln(L_\nu)/\partial\psi_r = 0$ for $r=0,\dots,c_\nu$,

where c_ν is the largest raw score $a_{vo}<k$ observed in subgroup G_ν. The solutions of Eqs. 21-22 are denoted by $\hat\sigma_1^{(\nu)},\dots,\hat\sigma_k^{(\nu)}$ and $\psi_0^{(\nu)},\dots,\psi_{c_\nu}^{(\nu)}$.

If the model holds, we should always estimate the same parameters $\sigma_i^{(\nu)} = \sigma_i$ and $\psi_r^{(\nu)} = \psi_r$, regardless of which subset of students is selected. This offers the possibility of checking the fit of the model through a comparison of the restricted CML estimates $(\hat\sigma_1^{(\nu)},\dots,\hat\sigma_k^{(\nu)}; \hat\psi_0^{(\nu)},\dots,\hat\psi_{c_\nu}^{(\nu)}$, $\nu=1,\dots,M$, with the unrestricted CML estimates $(\hat\sigma_1,\dots,\hat\sigma_k; \hat\psi_0,\dots,\hat\psi_{c_{max}})$.

Let $L(\sigma_1,\ldots,\sigma_k;\psi_0,\ldots,\psi_{c_{max}})$ be the conditional likelihood

$\prod_{v=1}^{n} p\{(a_{vi})|a_{vo}\}$ and let $L^{(\nu)}(\sigma_1,\ldots,\sigma_k;\psi_0,\ldots,\psi_{c_\nu})$ be the corresponding restricted conditional likelihood

$\prod_{v\in G_\nu} p\{(a_{vi})|a_{vo}\}$.

Then, we define a conditional likelihood ratio λ by

$$(23)\qquad \lambda = \frac{L(\hat\sigma_1,\ldots,\hat\sigma_k;\hat\psi_0,\ldots,\hat\psi_{c_{max}})}{\prod_{\nu=1}^{M} L^{(\nu)}(\hat\sigma_1^{(\nu)},\ldots,\hat\sigma_k^{(\nu)};\hat\psi_0^{(\nu)},\ldots,\hat\psi_{c_\nu}^{(\nu)})}\ .$$

Since $L(\sigma_1,\ldots,\sigma_k;\psi_0,\ldots,\psi_{c_{max}}) = \prod_{\nu=1}^{M} L^{(\nu)}(\sigma_1,\ldots,\sigma_k;\psi_0,\ldots,\psi_{c_\nu})$

and since $(\hat\sigma_1^{(\nu)},\ldots,\hat\sigma_k^{(\nu)};\psi_0^{(\nu)},\ldots\psi_{c_\nu}^{(\nu)})$ maximizes the ν-th factor in the denominator of Eq. 23, we find that $\lambda\leq 1$. If the model holds, then the restricted CML estimates should differ only slightly from the overall estimates. Values of λ close to 1 therefore indicate a good fit of the model; if λ is substantially smaller than 1, the model will be rejected.

From a theorem by ANDERSEN (1971), it follows that the distribution of $-2\ln(\lambda)$ converges for $n\to\infty$ to a χ^2-distribution with

$$df = (k-1)(M-1) + \sum_{\nu=1}^{M} c_\nu - c_{max}$$

degrees of freedom. The model will be rejected at an asymptotic significance level α when $-2\ln(\lambda)$ is larger than the $(1-\alpha)$ percentile of the limiting χ^2-distribution. As usual, the degrees of freedom are the number of free parameters specified under the hypothesis.

6. The Statistical Significance of Transfer Effects

In the microevaluation of training material we shall be interested to determine whether transfer or learning actually does occur, or whether the data allow for a reduction in the number of parameters, assuming that no learning or whatever form of transfer takes place

(24) $$\psi_0 = \psi_1 = \ldots = \psi_{c_{max}} = 0 \ .$$

Under the null hypothesis (Eq. 24), the conditional likelihood (Eq. 18) is reduced to

(25) $$L_0(\sigma_1,\ldots,\sigma_k|\psi_0=\ldots=\psi_{c_{max}}=0) = \frac{\prod_{i=1}^{k} \sigma_i^{n-a_{oi}}}{\prod_{s=1}^{k-1} \gamma_s(k)^{N_{k-s}}}$$

and we obtain CML estimates for the item difficulties by taking logarithms of Eq. 25 and setting the first order partial derivatives $\partial \ln(L_0)/\partial\sigma_\alpha = 0$, which yields the conditional estimation equations

(26) $$\frac{n-a_0}{\sigma_\alpha} - \sum_{s=1}^{k-1} N_{k-s} \frac{\gamma_{s-1}^{(\alpha)}(k)}{\gamma_s(k)} = 0 \quad \text{for } \alpha=1,\ldots,k \ .$$

For testing the null hypothesis we may then consider the likelihood ratio

(27) $$\lambda^* = \frac{L_0(\hat\sigma_1^{(0)},\ldots,\hat\sigma_k^{(0)}|\psi_0=\ldots=\psi_{c_{max}}=0)}{L(\hat\sigma_1,\ldots,\hat\sigma_k;\hat\psi_0,\ldots,\hat\psi_{c_{max}})} \ ,$$

where $\hat\sigma_1^{(0)},\ldots,\hat\sigma_k^{(0)}$ denote the solutions to Eqs. 26. The null hypothesis will be rejected when λ^* is substantially smaller than 1. Again, we make use of the theorem by ANDERSEN (1971) and approximate the distribution of $-2\ln(\lambda^*)$ by a χ^2-distribution with $df=c_{max}$ degrees of freedom. Hence, the null hypothesis can be rejected at an asymptotic significance level α

when $-2\ln(\lambda^*)$ is larger than the $(1-\alpha)$ percentile of the χ^2-distribution with $df=c_{max}$ degrees of freedom.

7. Parametric Hypothesis Testing: the Comparison of Alternative Training Materials

Once having determined that the training material is producing a transfer effect, the evaluator will be interested in testing hypotheses about the parametric structure: are alternative training items equally difficult? Does alternative training material produce the same kind of transfer? Does the transfer already approach some asymptotic level after a limited number of correct answers? etc..

Let us consider m groups of students and m corresponding sets of training materials. Then, we denote the number of items in the training material exposed to group G_j by k_j and write σ_{ij} for the difficulty of the i-th item to which the students from group G_j respond. Similarly, we write ψ_{rj} for the transfer effect produced by r "positive" responses to items of the training material exposed to group G_j. With each group of students we associate a response matrix $((a_{vi}^{(j)}))$ and a raw score vector $(a_{vo}^{(j)})$. The conditional likelihood of the data cube

$$(((a_{vi}))) = \begin{bmatrix} ((a_{vi}^{(1)})) \\ \vdots \\ ((a_{vi}^{(m)})) \end{bmatrix}$$

given the raw score matrix

$$((a_{vo})) = \begin{bmatrix} (a_{vo}^{(1)}) \\ \vdots \\ (a_{vo}^{(m)}) \end{bmatrix}$$

follows from

$$(28)\quad L(((\sigma_{ij})),(i=1,k_j;j=1,m);((\psi_{rj})),(r=0,c_j;j=1,m)) =$$

$$= p\,\{(((a_{vi}))) \mid ((a_{vo}))\} = \prod_{j=1}^{m} p\,\{((a_{vi}^{(j)})) \mid (a_{vo}^{(j)})\}$$

$$= \prod_{j=1}^{m} \frac{\prod_{i=1}^{k_j} \prod_{r=0}^{i-1} (\sigma_{ij} - \psi_{rj})^{n_{rij}}}{\prod_{s=1}^{k_j-1} \left\{ \sum_{m=0}^{s} \delta_m(k_j-s)\, \gamma_{s-m}(k_j)(-1)^m \right\}^{N_{k_j-s,j}}} .$$

Conditional maximum likelihood estimates $\hat{\sigma}_{1j},\ldots,\hat{\sigma}_{k_j j}$ and $\hat{\psi}_{0j},\ldots,\hat{\psi}_{c_j j}$ for the item and transfer parameters pertaining to group G_j are obtained from maximizing the j-th factor in the expression on the right hand side of Eq. 28. From a comparison of Eqs. 28 and 18 we see, that the estimation equations have the same form as in Eqs. 19-20.

Now let x and y be positive integers $x< \sum_{j=1}^{m} k_j$ and $y< \sum_{j=1}^{m} (c_j+1)$ We want to test whether the parameter space can be reduced to x item parameters and y transfer parameters. For instance, we might want to test the hypothesis that all items are equally difficult

$$\sigma_{ij} = \sigma \qquad \text{for } i=1,\ldots,k_j \text{ and } j=1,\ldots,m;$$

that the transfer is the same for all subgroups of students

$$\psi_{rj} = \psi_r \qquad \text{for } r=0,\ldots,c_j \text{ and } j=1,\ldots,m;$$

and that the transfer already approaches its asymptotical level after r=q correct answers

$$\psi_r = \psi_q \qquad \text{for } r=q,\ q+1,\ldots \quad .$$

Then we may introduce two sets of selection vectors $f_{ij} = (f_{ij}^{(1)},\ldots,f_{ij}^{(x)})$ and $g_{rj} = (g_{rj}^{(1)},\ldots g_{rj}^{(y)})$ and two sets of basic parameters $\eta_1,\ldots,\eta_x$ and $\Phi_1,\ldots,\Phi_y$. (A selection vector is a vector with all components equal to 0 exept one with the value 1. Accordingly, there are z possible selection vectors of dimension z $(1,0,0,\ldots,0)$, $(0,1,0,\ldots,0),\ldots,(0,0,\ldots,1)$.

With this notation, the hypothesis can be formulated in terms of specification equations

$$(29)\qquad \sigma_{ij} = \sum_{h=1}^{x} f_{ij}^{(h)} \eta_h \qquad \text{for } i = 1,\ldots,k_j \text{ and } j = 1,\ldots,m$$

and

$$(30)\qquad \psi_{rj} = \sum_{h=1}^{y} g_{rj}^{(h)} \Phi_h \qquad \text{for } r = 0,\ldots,c_j \text{ and } j = 1,\ldots,m$$

For the hypothesis in our example the selection vectors would be $f_{ij} = (f_{ij}^{(1)}=1)$ for $i=1,\ldots,k_j$ and $j=1,\ldots,m$; $g_{rj} = (0,\ldots,0,g_{rj}^{(r+1)}=1,0,\ldots,0)$ for $r\leq q$ and $j=1,\ldots,m$, and $g_{rj} = (0,\ldots,0,g_{rj}^{(q+1)}=1)$ for $r\geq q$ and $j=1,\ldots,m$.

The basic parameters are estimated by setting the first order partial derivatives of $\ln(L) = \Sigma_{j=1}^{m}\ln(L_j)$ equal to 0, where L_j denotes the j-th factor in the expression on the right hand side of Eq. 28:

$$(31)\qquad \frac{\partial \ln(L)}{\partial \eta_\alpha} = \sum_{j=1}^{m}\sum_{i=1}^{k} \frac{\partial \ln(L_j)}{\partial \sigma_{ij}} \frac{\partial \sigma_{ij}}{\partial \eta_\alpha} = \sum_{j=1}^{m}\sum_{i=1}^{k} \frac{\partial \ln(L_j)}{\partial \sigma_{ij}} f_{ij}^{(\alpha)} = 0 \text{ for } \alpha=1,\ldots,x$$

and

$$(32)\qquad \frac{\partial \ln(L)}{\partial \Phi_\beta} = \sum_{j=1}^{m}\sum_{r=0}^{c_j} \frac{\partial \ln(L_j)}{\partial \psi_{rj}} \frac{\partial \psi_{rj}}{\partial \Phi_\beta} = \sum_{j=1}^{m}\sum_{r=0}^{c_j} \frac{\partial \ln(L_j)}{\partial \psi_{rj}} g_{rj}^{(\beta)} = 0 \text{ for } \beta=1,\ldots,y\ .$$

Finally, we consider the likelihood ratio

$$(33)\qquad \lambda^{**} = \frac{L((\hat{\eta}_l),(l=1,\ldots,x);\ (\hat{\Phi}_h),(h=1,\ldots,y))}{L(((\hat{\sigma}_{ij})),(i=1,k_j;j=1,m);\ ((\hat{\psi}_{rj})),(r=0,c_j;j=1,m))}$$

for testing the hypothesis (Eqs. 29-30). Making use of the asymptotic results by ANDERSEN (1971), the hypothesis will be rejected at an asymptotic significance level α when $-2\ln(\lambda^{**})$ is larger than the $(1-\alpha)$ percentile of the limiting χ^2-distribution with

$$df = \sum_{j=1}^{m} (k_j+c_j-1) - (x+y-2)$$

degrees of freedom.

8. Student Evaluation

From Eq. 12 it follows that the likelihood of the data matrix $((a_{vi}))$ is dependent on the item difficulties via the item marginal matrix $((n_{ri}))$ only, so that the conditional probability

$$(34)\qquad p\{((a_{vi}))|((n_{ri}))\} = \frac{\prod_{v=1}^{n}\prod_{r=0}^{a_{vo}-1}(\xi_v+\psi_r)}{\sum_{((a^*_{vi}))|((n_{ri}))}\prod_{v=1}^{n}\prod_{r=0}^{a^*_{vo}-1}(\xi_v+\psi_r)}$$

does not contain the item parameters any more. In KEMPF (1974a) I have suggested that comparisons of the students' abilities could be based on Eq. 34. Since the conditional likelihood $p\{((a_{vi}))|(n_{ri}))\}$ cannot be used as a basis for parameter estimation, however, such comparisons have no practical relevance, but only interpretative meaning.[1]

1 Since a_{vo} is a sufficient statistic for the individual parameter ξ_v, the estimated ability parameters must be the same for all students who have the same raw score. Inserting the parameter estimates $\hat{\xi}_v=\hat{\xi}_u$ for $a_{vo}=u$ into Eq. 34 yields the expression

$$\prod_{u=1}^{k-1}\prod_{r=0}^{u-1}(\hat{\xi}_u+\psi_r)^{N_u} \Big/ \sum_{((a^*_{vi}))|((n_{ri}))}\prod_{u=1}^{k-1}\prod_{r=0}^{u-1}(\hat{\xi}_u+\psi_r)^{N^*_u},$$

in which the parameter estimates diminish, since the raw score frequencies are determined by the item marginals so that $N_u=N^*_u$ for $u=1,\dots,k-1$ and all possible response matrices $((a^*_{vi}))$ which are compatible with the item marginal matrix $((n_{ri}))$. An analoguous situation arises in estimating the individual parameters of the Rasch model when $k\leq 3$.

Let us consider the matrix of responses of two students with raw scores $a_{1o}=u_1$ and $a_{2o}=u_2$. Without loss of generality we may assume that $u_1<u_2$. Then, the expression on the right hand side of Eq. 34 takes the form

$$\frac{\prod_{r=0}^{u_1-1} (\xi_1 + \psi_r) \prod_{r=0}^{u_2-1} (\xi_2 + \psi_r)}{\prod_{r=0}^{u_1-1} (\xi_1 + \psi_r) \prod_{r=0}^{u_2-1} (\xi_2 + \psi_r) + \prod_{r=0}^{u_2-1} (\xi_1 + \psi_r) \prod_{r=0}^{u_1-1} (\xi_2+\psi_r)} =$$

$$= \frac{1}{1 + \prod_{r=u_1}^{u_2-1} \{(\xi_1 + \psi_r)/(\xi_2 + \psi_r)}$$

and thus yields a measure for the dissimilarity of the students' respective abilities $\xi_{vr} = \xi_v + \psi_r$ for $r=u_1,\ldots,u_2-1$.

Now let us assume that the transfer approaches its aymptotic level $\psi_\infty>\psi_0$ after q correct answers so that all students who were successful on q or more of the training items have the same learning gain $\Delta=\psi_\infty - \psi_0$. Whatever a student's initial ability $\xi_{v0} = \xi_v + \psi_0$ may be, he will then always have the possibility to reach the mastery criterion of q successful trials - and the corresponding learning gain Δ - if he only continues training long enough. Variations in the students' initial abilities may cause some students to have a better chance to reach the mastery criterion after a smaller number of trials than others. There will also still be a variation in the students' asymptotic abilities $\xi_{v\infty} = \xi_v + \psi_\infty$. As regards the probabilities of correct responses to future items, however, the learning gain Δ yields a steeper ascent for those students who had the lower initial ability (cf. Figure 3), and the dissimilarity in the students' actual achievement on the future items will be smaller than without prior training.

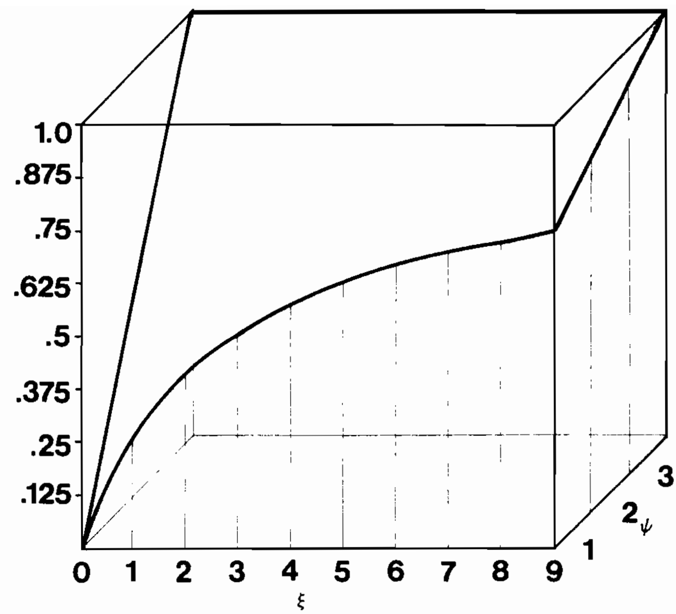

Figure No. 3.: The conditional item characteristic (Eq. 4) as a function of the latent ability variable ξ and of the transfer parameter ψ for fixed item difficulty σ =3.

If the above assumptions hold, the main purpose of student evaluation will be to determine whether or not a student has reached the mastery criterion and if not, to provide additional training material. For estimating students' asymptotic abilites, we may then apply the standard algorithms for estimating individual parameters of the Rasch model to responses $(a_{v\rho_v},\ldots,a_{vk})$, where ρ_v is the number of the first item to which student v responds after having reached the mastery criterion. On the basis of these estimates the teacher may then decide whether a further improvement of a student's ability is necessary. In order to achieve this, the teacher might resort to individualized teaching methods or the like.

References:

ANDERSEN, E.B.: The asymptotic distribution of conditional likelihood ration tests. Journal of the American Statistical Association 1971, 66, 630-633.

ANDERSEN, E.B.: Conditional inference and models for measuring. Copenhagen: Mentalhygiejnisk Forlag, 1973.

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. University of Vienna: Department of Psychology, 1972.

KEMPF, W.F.: Dynamische Modelle zur Messung sozialer Verhaltensdispositionen. In: KEMPF, W.F. (ed.): Probabilistische Modelle in der Sozialpsychologie. Bern: Huber, 1974a.

KEMPF, W.F.: A dynamic model for measuring individual and situative influences on social behavior. In COMPSTAT 74. Proceedings in Computational Statistics. Vienna: Physica, 1974b.

KEMPF, W.F.: Einige dynamische Meßmodelle und ihre Anwendung in der Denk- und Lernpsychologie. In: KEMPF, W.F. (ed.): Drei Arbeiten über verallgemeinerte Rasch-Modelle/Three papers on generalized Rasch-models. Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel: IPN-Arbeitsbericht 10, 1974c.

KEMPF, W.F.: A test-theoretical approach to structural learning. In: SCANDURA, J.M., DURNIN, J.H., WULFECK II, W.H. (eds.): Proceedings of the 5th annual interdisciplinary conference on structural learning. Philadelphia: MERGE Research Institute, Technical Report, 1974d, 310-322.

KEMPF, W.F.: Dynamic models for the measurement of "traits" in social behavior. In: KEMPF, W.F., & REPP, B.H. (eds.): Some mathematical models for social psychology. Bern: Huber, 1976a (in print).

KEMPF, W.F.: Ein dynamisches Testmodell mit additiven Nebenbedingungen bezüglich der Itemparameter. In: KEMPF, W.F., NIEHUSEN, B., MACH, G. (eds.): Logistische Testmodelle mit additiven Nebenbedingungen. Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel: IPN-Arbeitsbericht 22, 1976b.

KEMPF, W.F., & HAMPAPA, P.: The numerical solution of a set of conditional estimation equations arising in a dynamic test model. In: KEMPF, W.F., HAMPAPA, P., & MACH, G. (eds.): Conditional maximum likelihood estimation for a dynamic test model. Arbeitsbericht 13, Institute for Science Education at the University of Kiel, 1975.

KEMPF, W.F., & HAMPAPA, P.: Conditional inference for the dynamic test model. In: KEMPF, W.F., & REPP, B.H. (eds.):

Some mathematical models for social psychology. Bern: Huber, 1976 (in print).

KEMPF, W.F., & MACH, G.: A Fortran program for CML estimation in a dynamic test model. In: KEMPF, W.F., HAMPAPA, P., & MACH, G. (eds.): Conditional maximum likelihood estimation for a dynamic test model. Arbeitsbericht 13, Institute for Science Education at the University of Kiel, 1975.

KEMPF, W.F., NIEHUSEN, B.: Algorithmen und Programme für ein logistisches Testmodell mit additiven Nebenbedingungen bezüglich der Itemparameter. In: KEMPF, W.F., NIEHUSEN, B., & MACH, G. (eds.): Logistische Testmodelle mit additiven Nebenbedingungen. Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel: IPN-Arbeitsbericht 22, 1976.

NEYMANN, J., & SCOTT, E.L.: Consistent estimates based on partially consistent observations. Econometrica, 1948, 16, 1-32.

RASCH, G.: Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielson & Lydiche, 1960.

RASCH, G.: On general laws and the meaning of measurement in psychology. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1961, 5, 321-333.