Psychological Test and Assessment Modeling, Volume 53, 2011 (2), 183-191

Applying the LLTM for the determination of children's cognitive age-acceleration function

Klaus D. Kubinger[1]*, Christine Hohensinn,*
Stefana Holocher-Ertl & Nina Heuberger

Abstract

The paper uses Item Response Theory (IRT) for modeling and hypothesis testing children's cognitive age-acceleration function – within calibration and standardization of some intelligence test. For this, basically Fischer's *Linear logistic test model* (LLTM; Fischer, 1973, 2005) is applied. However, instead of originally decomposing the item difficulty parameters of the Rasch model into certain hypothesized elementary parameters, we now suggest to decompose the person parameter alike. That is, there is a decomposition into a testee's basic ability parameter and an age-leveled effect due to the developmental stage of the age-group in question. For convenience, we only interchange testees and items in order to facilitate parameter estimation and model test – of course, the Rasch model is totally symmetric as concerns testees and items. By doing so, all findings in the context of LLTM apply; in particular, pertinent program packages are at our disposal. In order to examine the suggested approach's feasibility, an empirical example is given. An Analogy test with eight items administered to more than 300 testees aged between 6 and 16, was analyzed. As a matter of fact, the logistic acceleration function proved to fit the data well and best.

Key words: Rasch model; LLTM; program package eRm; cognitive age-acceleration; Adaptive Intelligence Diagnosticum

[1] *Correspondence concerning this article should be addressed to:* Klaus D. Kubinger, PhD Prof., Head of the Division of Psychological Assessment and Applied Psychometrics, Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria, Europe; email: klaus.kubinger@univie.ac.at

Introduction

These days, psychological test calibration very often applies models of Item Response Theory (IRT). Above all, the Rasch model, or to say the 1-PL model, is one of the most used ones. However, the psychological test construction as a whole concerns classical test theory as well, that is in particular the test standardization. If for instance age-based differences in test scores occur, a specific standardization is carried out for every age-group. This is true for almost every developmental or intelligence test. However, in this manner, we cannot gain any hypothesis proven mathematical function of the given age-acceleration. At most, as for instance recently done by Kubinger (2009a), we can smooth some curvilinear function to the gradient of test scores' means in different age-groups. However it would be preferable to estimate the age-accelerating function by means of a specified model.

In the framework of IRT, several models have been already developed for modeling developmental growth or learning processes in the latent variable. Very recently von Davier, Xu, and Carstensen (2011) gave an elaborated overview, comparison, and application of latent growth models. Some of them rely on repeated measurements which is not appropriate for our setting of test calibration. For others (e.g the Saltus model by Wilson (1989)) no software implemention for incomplete data sets using CML (conditional maximum likelihood)-based parameter estimation is actually available. Therefore starting from a very practical point of view, we attempt to test different hypotheses of children`s cognitive age-acceleration function by using Fischer`s *Linear logistic test model* (LLTM; Fischer, 1973, 2005). For this approach CML-based software implementation is already available even for incomplete data sets.

Method

The Linear logistic test model (LLTM)

Recently, Kubinger (2008, 2009b) pointed out several LLTM applications which were not thought of for this model initially – bear in mind that originally (Fischer, 1973) some linear combination of certain hypothesized elementary parameters had been used to decompose the item parameter of the Rasch model (perhaps the most striking example is provided recently by Sonnleitner, 2008, where different item radicals have been disclosed, which more or less govern the difficulty of reading comprehension test items; see also Poinstingl, 2009, who established with the use of LLTM item generating rules for constructing a new type of reasoning items with difficulties determined by the examiner). LLTM's other applications deal with: a) Rasch model item calibration using data sampled consecutively in date but partly from the same testees; b) measuring position effects of item presentation, in particular, learning and fatigue effects – specific for each position, as well as linear or non-linear; c) measuring content-specific learning effects; d) measuring warming-up effects; e) measuring effects of speeded item presentation; f) measuring effects of different item response formats. In particular, b) and f) have been applied several times since then (e.g. Hohensinn, Kubinger, Reif, Holocher-Ertl, Khorramdel, & Frebort, 2008; Hohensinn & Kubinger, 2011).

We will illustrate a new approach of using LLTM. That is, we swap testees and items within the Rasch model data design. This with the aim to decompose the person parameter into at least two component parameters instead of decomposing the item parameters into some linear combination of operation parameters.

Bear in mind, that the LLTM originally defines the probability of a testee *v* with the ability parameter ξ_v of solving item *i* with the Rasch model difficulty parameter σ_i as follows:

$$P\left(+\middle|\xi_v;\sigma_i=\sum_j^p q_{ij}\eta_j\right)=\frac{e^{\xi_v-\sum_j^p q_{ij}\eta_j}}{1+e^{\xi_v-\sum_j^p q_{ij}\eta_j}} \tag{1}$$

– the parameters η_j ($j = 1, 2, \dots p < k$) are hypothesized elementary operation parameters which constitute the item parameter σ_i, $i = 1, 2, \dots k$; the values q_{ij} are postulated as being fixed and known weights. However, if one is interested in individual ability parameters $\xi_{v(g)}$ whereby $v(g)$ indicates that testee *v* stems from age-group *g*, we might think of some decomposition as $\xi_{v(g)} = \xi^*_v + \lambda_g$, that is a decomposition into a testee's basic ability parameter ξ^*_v and an age-leveled effect λ_g due to the developmental stage of the age-group in question. Formula (1) then changes to

$$P\left(+\middle|\xi_{v(g)}=\xi_v^*+\lambda_g;\sigma_i\right)=\frac{e^{\xi_v^*+\lambda_g-\sigma_i}}{1+e^{\xi_v^*+\lambda_g-\sigma_i}} \tag{2}$$

Now we could derive conditional parameter estimation functions for the person parameters $\xi_{v(g)}$ as Rasch (1960/1980) did for item parameters σ_i and for the person component parameters ξ^*_v and λ_g as Fischer (1973) did for the elementary operation parameters η_j, respectively. And we could program some software for the respective parameter estimation as in particular the program package *eRm* (Mair, Hatzinger, & Maier, 2010; cf. also Poinstingl, Mair & Hatzinger, 2007) offers. Moreover, we would have to adapt Andersen's Likelihood ratio test in order to check whether the Rasch model holds and then to adapt the goodness-of-fit test to see whether the data's likelihood according to model (2), L^*, fits the data's likelihood according to the Rasch model – with regard to LLTM it is well-known that the data's likelihood according to model (1), L_{LLTM}, is to oppose to the data's likelihood according to the Rasch model, L_{RM}, whereby $-2\ln(L_{\text{LLTM}}/L_{\text{RM}})$ is asymptotically χ^2-distributed with $df = k - p - 1$.

However, we will not do so. Instead, we only interchange testees and items. Indeed, the Rasch model is completely symmetric regarding testees and items (cf. Rasch, 1960/1980). Though, in the first instant, the decomposition of persons' or items' parameters is not symmetric; that is true as long as conditional parameter estimation with respect to the item parameters is aimed for, which is the case in LLTM. But in the second instant, if we actually interchange testees and items for the application of LLTM as well, decomposition of the person parameters works.

That is, from now on we define the person parameters of k testees as σ_i, i = 1, 2, … k. Then we define the item parameters of n items as ξ_v , v = 1, 2, … n. The only unusual thing is that we then have a larger number of items n than testees k. Rasch model analysis might apply in a state of the art manner. The same is true concerning Andersen's Likelihood ratio test and graphical model checks. For the former the asymptotic property might be examined, at most. Regarding LLTM, formula (2) now turns into formula (3); thereby decomposing the individual ability parameters $\sigma_{i(g)}$ (i.e. testee i stems from age-group g) into $\sigma_{i(g)} = \sigma^*_i + \lambda_g$:

$$P\left(+\middle|\xi_v;\sigma_{i(g)}=\sigma_i^*+\lambda_g\right)=\frac{e^{\xi_v-(\sigma_i^*+\lambda_g)}}{1+e^{\xi_v-(\sigma_i^*+\lambda_g)}} \tag{3}$$

However, in order to apply model (3) the data have to be re-arranged, or to say, the obligatory LLTM structure matrix has to be well defined. Given the application of the Rasch model, the structure matrix in terms of formula (1) with testees and items interchanged appears as shown in Figure 1. For each testee i a specific person parameter σ_i is to be estimated.

For formula (3) however the structure matrix appears as shown in Figure 2. Given the Rasch model holds, there are different score groups, that is groups of testees having the same number of solved items and as a consequence of which, they have the same person parameter (estimation). As the number of solved items varies between r = 1 and $r = n – 1$ (r = 0 and $r = n$ deliver no information), these parameters are $\sigma^{(r)}$. However, as the same number of solved items might occur for testees being of age g as well as being of age h, these parameters have to be decomposed, as already indicated into a basic parameter $\sigma^{*(r)}$

testee i	1	2	3	4	5	6	…	k
person parameter i	σ_1	σ_2	σ_2	σ_4	σ_5	σ_6	σ_i	σ_k
1	1							
2		1						
3			1					
4				1				
5					1			
6						1		
…							1	
k								1

Figure 1:
The structure matrix $((q_{ij}))$ of the Rasch model (an empty cell means $q_{ij} = 0$)

component	parameter / testee $i(r, g)$	1(1, 0)	…	i’(1, 1)	…	i’’(1, g)	…	…	i’’’(5, g)	…	k(r, g)
1	$\sigma^{*(1)}$	1		1		1					
2	$\sigma^{*(2)}$										
3	$\sigma^{*(3)}$										
4	$\sigma^{*(4)}$										
5	$\sigma^{*(5)}$								1		
…											
r	$\sigma^{*(r)}$										1
…											
n-1	$\sigma^{*(n-1)}$										
…	λ_0	1									
	λ_1			1							
	…										
	λ_g					1			1		1
	…										

Figure 2:
The structure matrix $((q_{ij}))$ of model (3a) (an empty cell means $q_{ij} = 0$)

and an age-leveled effect λ_g due to the developmental stage of the age-group g. That is formula (3) has to be specified as follows, because testee i is to index according to the score group – in terms of the formula (1) $\eta_j = \sigma^{*(r)}$ ($j = r = 1, 2, \dots n - 1 < k$), $\eta_{n-1+g} = \lambda_g$ ($g = 0, 1, 2, \dots$); $q_{ij} = q_{i(r)j+g}$ is either one or zero:

$$P\left(+\middle|\xi_v;\sigma_{i(r,g)} = \sigma^{*(r)} + \lambda_g\right) = \frac{e^{\xi_v - (\sigma^{*(r)} + \lambda_g)}}{1 + e^{\xi_v - (\sigma^{*(r)} + \lambda_g)}} \qquad (3a)$$

Of course, the data matrix is as usual, again just testees and items were interchanged (cf. Figure 3).

As in the original LLTM the structure matrix has to be standardized to an “anchor”, for instance this is set to $\sigma^{*(n-1)} = \lambda_0 = 0$ here; otherwise the matrix would not have full rank and the estimations would not be indeterminate.

items / testee *i*(*r*, *g*)	1(1, 0)	…	*i’*(1, 1)	…	*i’’*(1, g)	…	…	*i’’’*(5, g)	…	*k*(*r*, *g*)
1	0		1		0			1		
2	1		0		0			0		
3	0		0		1			1		
4	0		0		0			1		
5	0		0		0			1		
…	0		0		0			0		
n	0		0		0			1		

Figure 3:
The data matrix for Rasch model and LLTM analyses, given testees and items were interchanged (in the given example testee 1 has solved only the second item, testee *i’* only the first, and testee *i’’* only the third. Testee *i’’’* has solved 5 items in total, number 1, 3, 4, 5, and *n*).

An empirical example

The data of a Rasch model-calibrated test were at our disposal, that is a new subtest, Analogies, of the next generation of the intelligence test-battery for children AID 2 (*Adaptive Intelligence Diagnosticum*; Kubinger, 2009a). There are $n = 8$ items which were administered to each of 348 testees, aged between 6 and 16. After the deletion of all testees with either $r = 0$ or $r = n = 8$ solved items $k = 308$ testees remained.

First of all, the subset of the 8 items in question was to be tested as to whether they fit the Rasch model. Andersen’s Likelihood ratio test according to the partition of the items which were solved rather often and those solved rather seldom, resulted in a non-significant test-statistic of $\chi^2 = 4.17 < \chi^2_{0.05} = 14.07$, $df = 7$. Hence the application of our suggested approach is justified.

Applying the LLTM as worked out above for the determination of children’s cognitive age-acceleration function, we aimed to test three hypotheses: 1) There is only a need for a single parameter λ which is, due to the age-groups 6, 7, … 15, weighted by 0, 1, 2, 3, 4, and so forth – that is, a linear acceleration function is supposed. 2) Again, there is only a need of a single parameter λ which is, due to the age-groups 6, 7, … 15, weighted by .50, .73, .88, .95, .98, .99, 1.00, 1.00, 1.00, 1.00 – that is a logistic acceleration function is supposed. 3) There are age-specific parameters λ_g as given in Figure 2 – that is no certain mathematical function is supposed.

Results

The Likelihood ratio tests according to the three hypotheses, that is $-2\ln(L_{LLTM} / L_{RM})$ with $df = k - [(n - 2) + 1] - 1 = k - n = 300$ in both the first cases and $df = k - [(n - 2) + (10 -$

1)] – 1 = $k - n - 8 = 292$ in the third case, resulted in: $\chi^2 = 6.30 < \chi^2_{0.05} = 341.40$, $\chi^2 = 0.72 < \chi^2_{0.05}$, and $\chi^2 = 16.07 < \chi^2_{0.05} = 332.85$. Descriptively speaken, hypothesis 1), the linear age-acceleration model, fits the data well, though hypothesis 2), the logistic age-acceleration model does fit better – as both hypotheses have the same number of degrees of freedom we cannot apply a significance test, in order to compare these models. However, we actually can apply a significance test, as it concerns a comparison of both the models according to hypothesis 2) and 3), which hold as well. As a result we obtain $\chi^2 = 15.34 < \chi^2_{0.05} = 16.92$ ($df = 9$). That is, there is, at least, no need for age-specific parameters λ_g. Furthermore, we now have to examine, whether there is a significant age-acceleration effect at all. In order to do this, there are two possible approaches. A rather convenient one is to examine only the confidence interval which is based on the estimated parameter $\hat{\lambda}$; regarding hypothesis and model 2), respectively, $\hat{\lambda}$ amounts to 0.510 and the 95%-confidence interval to [–0.914; 1.934]; because 0 lies within this interval, the acceleration effect λ can even be 0 – the same is true with regard to hypothesis and model 1), respectively: $\hat{\lambda} = 0.416$, [–0.156; 0.988]). Some IRT-based approach is however to hypothesize a fourth model: 4) There is no need for a parameter λ. The respective likelihood is to be compared with that of hypothesis and model 2). As a matter of fact, that likelihood ratio test results in $\chi^2 = 18.80 > \chi^2_{0.05} = 3.84$ ($df = 1$). In this way, hence, the logistic age-acceleration effect λ has proven to be significant – though it is still, as indicated, rather small. At any rate, we have at least established a logistic trend of cognitive age-acceleration. Given the estimates of the basic parameter $\sigma^{*(r)}$ for testees with $r = 1, 2, \ldots 7$ solved items as –2.790; –1.732; –.830; –.002; 0.816; 1.725; 2.812, then there occurs a bonus effect by age, that is for instance with 8-year old children –2.790 + .88·0.510 = –2.341 and so on.

Discussion

From the point of view of users, we are of the opinion that our approach is currently the most convenient for the standardization of a Rasch model-calibrated psychological test. That is, we remain within the frame of reference of CML-based IRT but do not have to use classical test theory methods or MML (marginal maximum likelihood)-based methods, which, incidentally, always depend on the actual selected sample – bear in mind that in our approach, in a certain age-group eventually over- or underrepresented testees with a score of $r = 0$ and $r = n$ are not taken into account. Moreover, applying our approach enables the researcher to plan forthcoming standardization studies more economically, that is, instead of sampling each of all age-groups, just a few of them would suffice; standardization for the missing age-groups could be done according to the established acceleration function.

Indeed, we have here experienced with the suggested approach for a single example only. All above, simulations studies are needed in order to analyze the Likelihood ratio tests' degree of accuracy concerning their χ^2-distribution approximation – recently Hohensinn et al. (2008) showed that the traditional tests of LLTM sometimes do not detect relevant effects because of a type-II-risk being too high. In particular, there is hardly any knowledge of problems regarding parameter estimation, given respective disproportions of the number of items and the number of testees. Hence, especially further research on this topic is required.[2]

References

Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fischer, G.H. (2005). Linear logistic test models. In *Encyclopedia of Social Measurement, 2,* 505-514.

Hohensinn, C. & Kubinger, K.D. (2011, in print). Applying Item Response theory methods to examine the impact of different response formats. *Educational and Psychological Measurement*.

Hohensinn, C., Kubinger, K.D., Reif, M., Holocher-Ertl, S., Khorramdel, L. & Frebort, M. (2008). Examining item-position effects in large-scale assessment using the Linear Logistic Test Model. *Psychology Science Quarterly, 50,* 391-402.

Kubinger, K.D. (2008). On the revival of the Rasch model-based LLTM: From composing tests by item generating rules to measuring item administration effects. *Psychology Science Quarterly, 50,* 311-327.

Kubinger, K.D. (2009a). *Adaptives Intelligenz Diagnostikum – Version 2.2 (AID 2) samt AID 2-Türkisch* [Adaptive Intelligence Diagnostum, AID 2-Turkey included]. Göttingen: Beltz.

Kubinger, K.D. (2009b). Applications of the Linear Logistic Test Model in Psychometric Research*. Educational and Psychological Measurement, 69,* 232-244.

Mair, P., & Hatzinger, R., Maier, M. (2010). eRm: extended Rasch modeling. R package version 0.13-0: http://cran.r-project.org/web/packages/eRm/.

Poinstingl, H. (2009). The Linear Logistic Test Model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test. *Psychology Science Quarterly, 51,* 123-134.

Poinstingl, H., Mair, P. & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling). Anwendung des Rasch-Modells (1-PL Modell) – Deutsche Version* [Manual of eRm. To apply the Rasch model – German Version]. Lengerich: Pabst.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Sonnleitner, P. (2008). Using the LLTM to evaluate an item-generating system for reading Comprehension. *Psychology Science Quarterly, 50,* 345-362.

---

[2] As a matter of fact, we realized in first simulation studies as well as in the given empirical example, that computational problems in the convergency of parameter estimation arise if the number of testees is much larger than 300 (and the number of items is not up to 30). Therefore, in the illustrated example we used a random sample comprising only half the size of the originally given data.

Wilson, M. (1989). Saltus: a psychometric model for discontinuity in cognitive development. *Psychological Bulletin, 105*, 276-289.

Von Davier, M., Xu, X. & Carstensen, C. H. (2011). Measuring growth in a longitudinal large-scale assessment with a general latent variable model. *Psychometrika, 76,* 318-336.