Measurement and Research Department Reports 2010-3

The Equivalence of LLTMs Depends on the Design

Timo M. Bechger
Gunter Maris
Huub H.F.M. Verstralen

Measurement and Research Department Reports 2010-3

The Equivalence of LLTMs Depends on the Design

Timo M. Bechger, Cito

Gunter Maris, Cito & University of Amsterdam

Huub H.F.M. Verstralen, Cito

Cito
Arnhem, 2010

Abstract

Previous work by Bechger, Verstralen, and Verhelst (2002) and Fischer (2004) on the equivalence of linear logistic test models implicitely assumed complete data. This note describes how identifiability and equivalence depends on the design, using a invented example.

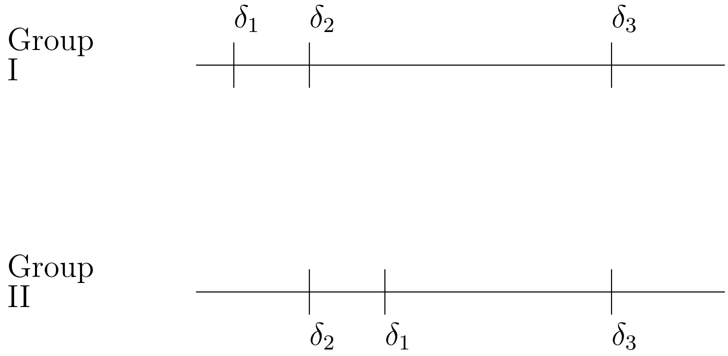

Figure 1.

Results of separate analysis. The positions of the items are their difficulties δ_i.

1. Introduction

The *linear logistic test model* (LLTM: see Fischer, 2007, par. 7, and references therein) has been developed as a model to test cognitive theories about the difficulties of the items. Bechger et al. (2002) describe how LLTMs that represent different cognitive theories may nevertheless be *equivalent* in the sense that they reproduce the same distribution. In this note we demonstrate that the equivalence of LLTMs depends on the design: a possibility that was overlooked in the previous work. This aim, we consider a simple example involving a non-equivalent groups, anchor test design.

2. Motivating Example

The same test of three items has been taken by two non-equivalent groups of children. Separate analyses of the data collected in each group show that items follow the Rasch model (Rasch, 1960). However, the relative difficulties of the items are *not* the same for both groups. As illustrated in Figure 1, items 1 and 2 have changed positions relative to one another, while the relative position of item 2 and item 3 is the same in both groups. Thus, the characteristics of the items depend on the group in which they are administered and we recognize this as a simple case of *Differential Item Functioning (DIF)*.

Next, we wish to do a concurrent analysis and analyze data from both groups simultaneously. To this aim, we need a set of *anchor items* to establish a relation between the two ability scales. Items that show DIF are kept in the analysis but removed from the anchor. They are treated as different items in the two populations. It will be clear that items 1 and 2 (or items 1 and 3) can not be both in the anchor but there is still a number of alternatives to choose from. We consider the following two possibilities, illustrated in Figure 2.

A1: Item 2 and item 3

A2: Item 1.

When we do the concurrent analyses, we find that both models give to the same likelihood, which suggests that they equivalent. Furthermore, the relation between the ability scales appears to be different for the two analyses. Consider a pupil in group II with an ability at the location of the second item, i.e., a pupil with a 50 percent chance to answer the second item correctly. In Figure 2 it is seen that, in the first analysis, the ability of this pupil would match the ability of a pupil in group I with a 50 percent chance to answer the second item correctly. However, this would no longer hold true for the second analysis.

3. Explanation

In the example, we stated that analyses with two different anchors lead to the same likelihood and the models are observationally equivalent models. At first sight, however, the model *cannot* be equivalent because they differ in the number of parameters. To wit, four parameters in the first analysis, and five in the second analysis

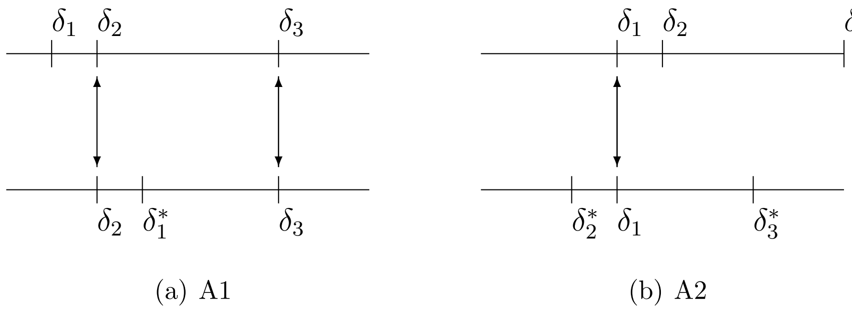

FIGURE 2.

Two different choices to link the two scales. An asterix indicates that the item is considered a different item in the second group.

(i.e., $\delta_1, \delta_2, \delta_3, \delta_2^*, \delta_3^*$). Consequently, one expects the second analysis to furnish a better fit to the data. This requires clarification.

On closer look, the difference between the two analyses is that, in the second analysis, there is no restriction that $\delta_1 - \delta_2 = \delta_2^* - \delta_3^*$. Thus, the second model is nested in the first and it is only *after* we impose the restriction that the two models become observationally equivalent. If we do *not* impose the restriction, it is possible to find that $\hat{\delta}_1 - \hat{\delta}_2 \neq \hat{\delta}_2^* - \hat{\delta}_3^*$ in which case one would decide in favour of the second model. When, however, $\delta_1 - \delta_2 = \delta_2^* - \delta_3^*$, this will become increasingly unlikely when sample size increases. To illustrate this, we have conducted a small simulation study. For different sample sizes, n, we generated 200 data sets and analyze each using *conditional maximum likelihood*. If we gather the conditional likelihoods for each model we find that the likelihood for the second analysis tends to be higher. However, as seen in Figure 3, the ratio of the likelihoods converges to one when sample size increases. In the sequel we will demonstrate this formally by casting the two models for concurrent analysis into the framework of the LLTM.

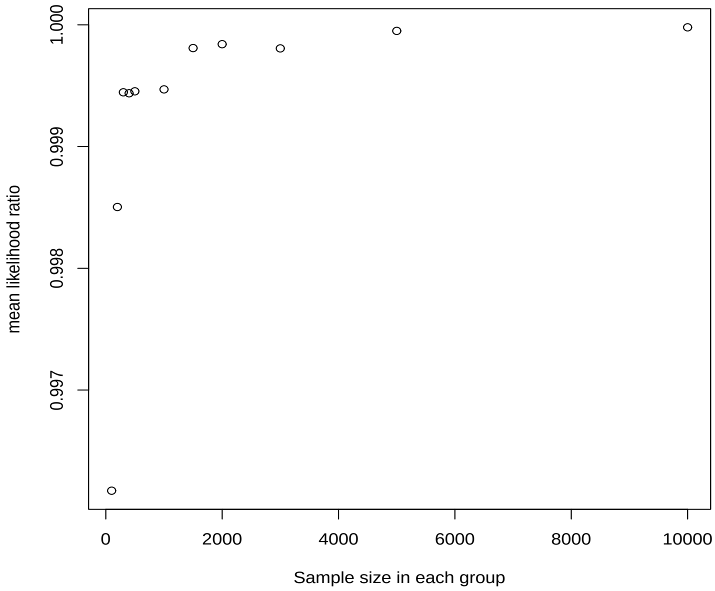

Figure 3.

Average Likelihood Ratio of the two concurrent analyses in the motivating example.

4. Two LLTMs

To impose the restriction $\delta_1 - \delta_2 = \delta_2^* - \delta_3^*$ we introduce a uniform translation or "shift" τ in the location of the DIF items between the two populations. That is, $\delta_i^* = \delta_i - \tau$. Thus, we have δ_1, δ_2, and δ_3, and a fourth parameter τ: $\delta_2^* = \delta_2 - \tau$ and $\delta_3^* = \delta_3 - \tau$. The model for the concurrent analysis is a Rasch model with a linear restriction on the parameters and a special case of the LLTM.

The LLTM is a Rasch model with linear restrictions on the item parameters. Under the Rasch model, the probability that person p finds the correct answer to item i is

$$P(X_{pi} = 1|\theta_p, \boldsymbol{\eta}) = \frac{\exp(\theta_p - \delta_i)}{1 + \exp(\theta_p - \delta_i)}.$$

In the LLTM, $\delta_i = \sum_j q_{ij}\eta_j$, where the q_{ij} are constants and the η_j are called *basic* *parameters*. The matrix containing the constants is traditionally called the **Q**-*matrix*. It is well-known that the Rasch model is not *identifiable*: I.e., an arbitrary constant can be added to δ_i and θ_p without changing the probability of a correct answer. This is solved with one identifying restriction called a *normalization*. We will also have to deal with this when we use the LLTM.

We will now specify the LLTMs corresponding to the two concurrent analyses that we did in the motivating example. To this aim, we define six "virtual" items with parameters $\delta_{i,j}$ that are linear functions of three basic parameters: η_1, η_2, and τ (see Fischer, 2007, pp. 568-571). The following LLTMs correspond to A1 and A2:

$$\begin{pmatrix} \delta_{1,1} \\ \delta_{2,1} \\ \delta_{3,1} \\ \delta_{1,2} \\ \delta_{2,2} \\ \delta_{3,2} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \tau \end{pmatrix} + c\mathbf{1} = \begin{pmatrix} -\eta_1 + c \\ c \\ \eta_2 - \eta_1 + c \\ \tau - \eta_1 + c \\ c \\ \eta_2 - \eta_1 + c \end{pmatrix} \tag{1}$$

$$\begin{pmatrix} \delta_{1,1} \\ \delta_{2,1} \\ \delta_{3,1} \\ \delta_{1,2} \\ \delta_{2,2} \\ \delta_{3,2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \tau \end{pmatrix} + c^*\mathbf{1} = \begin{pmatrix} c^* \\ \eta_1 + c^* \\ \eta_2 + c^* \\ c^* \\ \eta_1 - \tau + c^* \\ \eta_2 - \tau + c^* \end{pmatrix}, \tag{2}$$

where $\delta_{i,j}$ denotes the parameter of the ith item in population j, and $\mathbf{1}$ a column vector of ones. The models are formulated in such a way that the basic parameters have the same *statistical* interpretation in the two models: i.e., $\eta_1 = \delta_{2,1} - \delta_{1,1}$, $\eta_2 = \delta_{3,1} - \delta_{1,1}$ and $\tau = \delta_{3,1} - \delta_{1,1} - (\delta_{3,2} - \delta_{1,2})$. Conform the way the motivating example was constructed, the relative difficulties of items in the same booklet (i.e., $\delta_{i,g} - \delta_{j,g}$) are the same in each analysis. Since the basic parameters are one-to-one functions of relative difficulties, they do not need a subscript for group. Note further that we have specified the models in such a way that a normalization is imposed: i.e., the location of one of the anchor items is fixed to zéro. The constants c and c^* are used to accommodate arbitrary changes in the normalization. These constants express the fact that the item parameters are defined only up to an arbitrary shift (see Fischer, 2007, p. 564).

The motivating example suggests that these two LLTMs are *statistically equivalent*: That is, they "look" different but reproduce the same distribution. The question is: Can we prove this? It is easy to see that two LLTMs are equivalent when, for any value of the basic parameters (η_1, η_2, τ), the item parameters (i.e., the δ's) are the same up to an additive constant. That is, when the models differ only in the normalization constant: i.e., $c \neq c^*$. In this case, however, it is not that simple because the positions of items administered in the second group relative to those administered in the first group is different for the two models. Specifically,

$$\boldsymbol{\delta}_{\text{Model 1}} - \boldsymbol{\delta}_{\text{Model 2}} = \begin{pmatrix} -\eta_1 \\ -\eta_1 \\ -\eta_1 \\ -\eta_1 + \tau \\ -\eta_1 + \tau \\ -\eta_1 + \tau \end{pmatrix} + (c - c^*)\mathbf{1}$$

Again, the situation is more complex than it seems at first. That the models may still be equivalent is a consequence of the design. In a design with equivalent groups, the p-values (i.e., the observed percentage correct answers for each item) are directly comparable across the groups. Hence, the models can not be equivalent, in this case, because they imply different p-values for the same items in different groups. This is different with *non*-equivalent groups where the p-values are not directly comparable. Here, there is nothing to prevent p-values from being different in different populations, and τ can be absorbed in the average difference between the ability distributions.

It is the burden of the ensuing sections to investigate this matter in more detail. First, we apply results from existing work on equivalence of LLTMs only to find out that it doesn't apply. Then, we demonstrate that the equivalence of the model follows when we take into account that the design is incomplete.

5. Apply Earlier Work

Earlier work focussed on the mapping from basic parameters $\boldsymbol{\eta}$ to suitably normalized item-parameters $\boldsymbol{\delta}$. Here, we look directly at the relation between $\boldsymbol{\eta}$ and the log-odds:

$$\rho_{pi} = \ln \frac{P(X_{pi} = 1)}{P(X_{pi} = 0)}.$$

First, note that the log-odds for person p are a linear function of his or her ability and the basic parameters. In matrix notation:

$$\boldsymbol{\rho}_{p.} = [\mathbf{I}, \mathbf{1}] \begin{pmatrix} -\boldsymbol{\delta} \\ \theta_p \end{pmatrix} \tag{3}$$

$$= [\mathbf{Q}, \mathbf{1}] \begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \end{pmatrix}. \tag{4}$$

This linear function is one-to-one, and the model parameters, $(\boldsymbol{\eta}, \theta_p)$ are identifiable, *iff* the augmented matrix $\mathbf{Q}^+ = [\mathbf{Q}, \mathbf{1}]$ is of full column rank: a result that was established by Fischer (1983). This is found to be true for both models (1) and (2). Hence, both models are identifiable. As an aside, we note from Equation 3 that

$(\boldsymbol{\delta}, \theta_p)$ is not identifiable, as mentioned earlier: The matrix $[\mathbf{I}, \mathbf{1}]$ is not of full column rank.

If (and only if) $\mathbf{Q}^+$ is in the column-space of $\mathbf{W}^+ = [\mathbf{W}, \mathbf{1}]$, does there exists a matrix[1] $\mathbf{A}$ such that $\mathbf{W}^+\mathbf{A} = \mathbf{Q}^+$ (Pringle & Rayner, 1971, pp. 9-10), and

$$\begin{aligned} \boldsymbol{\rho}_{p.} &= \mathbf{Q}^+ \begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \end{pmatrix} \\ &= \mathbf{W}^+\mathbf{A} \begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \end{pmatrix} \\ &= \mathbf{W}^+ \begin{pmatrix} -\boldsymbol{\eta}^* \\ \theta_p^* \end{pmatrix}. \end{aligned} \tag{5}$$

Hence, the model defined by $\mathbf{Q}$ is *nested* in the model defined by $\mathbf{W}$: That is, given $(\boldsymbol{\eta}, \theta_p)$ we can always find $(\boldsymbol{\eta}^*, \theta_p^*)$ such that Equation 5 holds. For example, if $\mathbf{W} = \mathbf{I}$, we find that any LLTM is a special case of the (un-normalized) Rasch model. In a similar way we can derive the condition under which the model defined by $\mathbf{W}$ is nested in the model defined by $\mathbf{Q}$.

Example 1 (Nested LLTMs)**.** *Consider,*

$$\begin{aligned} \mathbf{Q}^+ &= \mathbf{W}^+\mathbf{A} \\ &= \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right) \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 3 & 0 & 1 \\ 1 & 0 & 1 \end{array}\right) \end{aligned}$$

[1] Actually, $\mathbf{A} = (\mathbf{W}^+)^{g_1}\, \mathbf{Q}^+$, where $(\mathbf{W}^+)^{g_1}$ denotes a *one-condition generalized inverse* of $\mathbf{W}^+$: i.e., any matrix $\mathbf{G}$ such that $\mathbf{W}^+\mathbf{G}\mathbf{W}^+ = \mathbf{W}^+$.

It follows that

$$\boldsymbol{\eta}^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_1 \end{pmatrix}$$

Hence, the LLTM defined by $\mathbf{Q}$ *is equivalent to the LLTM defined by* $\mathbf{W}$ *when* $\eta_1^* = \eta_3^*$. *Thus, the model defined by* $\mathbf{Q}$ *is a special case of the model defined by* $\mathbf{W}$. *The matrix* $\mathbf{A}$ *is singular and it is easily checked that there is no matrix* $\mathbf{B}$ *such that* $\mathbf{W}^+ = \mathbf{Q}^+\mathbf{B}$. *Hence, given* $(\boldsymbol{\eta}^*, \theta_p^*)$ *there is no* $(\boldsymbol{\eta}, \theta_p)$ *such that Equation 5 holds except when* $\eta_1^* = \eta_3^*$.

Two LLTMs are equivalent when each model is nested in the other. This is true *iff* $\mathbf{Q}^+$ and $\mathbf{W}^+$ span the same column space (Bechger et al., 2002, Theorem 2). It is easily proven that this is equivalent to the *non*-singularity[2] of $\mathbf{A}$ which shows that equivalent models must have the same number of basic parameters. The following proposition provides a simple, necessary and sufficient, condition for equivalence:

Proposition 2. *Two identifiable LLTMs with the same number of basic parameters and weight matrices* $\mathbf{Q}$ *and* $\mathbf{W}$ *are equivalent* iff *the rank of the augmented matrix* $[\mathbf{Q}^+, \mathbf{W}^+]$ *equals one plus the number of basic parameters in each model.*

Proof. See Appendix □

Note that the condition, in Proposition 2, that the LLTMs be identified is necessary because otherwise an LLTM might not be equivalent to itself (cf. Fischer, 2004, p. 311). The following example illustrates how Proposition 2 can be applied.

[2] $\mathbf{A}$ is non-singular if it is a square invertable matrix

Example 3 (Equivalent LLTMs)**.** *Consider the following design matrices:*

$$\mathbf{Q} = \begin{pmatrix} \mathit{0} & \mathit{0} \\ \mathit{-1} & \mathit{0} \\ \mathit{-1} & \mathit{1} \end{pmatrix} \quad \textit{and} \quad \mathbf{W} = \begin{pmatrix} \mathit{1} & \mathit{0} \\ \mathit{0} & \mathit{0} \\ \mathit{0} & \mathit{1} \end{pmatrix}$$

It is easily checked that the rank of $[\mathbf{Q}^+, \mathbf{W}^+]$ *equals* 3*. It follows from Proposition 2 that the models are equivalent. In fact, both models are equivalent to the Rasch model. The number of basic parameters equals the number of relative difficulties and if we look at, say, differences with respect to the difficulty of the first item (i.e.,* $\delta_i - \delta_1$*) we find that they are equal under both models. The models differ only in the normalization that they imply. In the first model, the first item parameter is set to zéro and in the second model it is the second item parameter.*

If we apply Proposition 2 to the two models (Eqs. 1 and 2) under consideration, we find that the matrix $[\mathbf{Q}^+, \mathbf{W}^+]$ is of rank 5, and not 4 as we expected when the models are equivalent. This suggests that either the models are *not* equivalent, or Proposition 2 is wrong. In what follows, we explain that by considering the log-odds of only one person, we have implicitly assumed that the data are complete. Proposition 2 applies in this case. However, a non-equivalent groups design yields *in*complete data in the sense that persons did not respond to all *virtual* items. Hence, Proposition 2 is not wrong: it simply doesn't apply here. The conclusion that the models under consideration are equivalent follows when we take proper account of the design.

6. Accounting for the Design

Consider the situation where the *same* people respond to the same items on *two* occasions. Assume that, on the second occasion, the ability of all students has changed with an amount τ. Hence, for a person p, the log-odds on the two occasions are:

$$\rho_{pi1} = \theta_p - \delta_i \tag{6}$$

$$\rho_{pi2} = \theta_p - \delta_i - \tau, \tag{7}$$

where $\tau = \rho_{pi2} - \rho_{pi1}, \forall\{i, p\}$ is identifiable, eventhough the person and item parameters are determined up to an additive constant. This hypothesis corresponds to a variant of the LLTM sometimes called the *saltus model* (Wilson, 1989; see also Fischer, 2000). Under this model, Equation 4 gives the log-odds for a single person. The log-odds for any two persons p and h can be written as:

$$\begin{pmatrix} \boldsymbol{\rho}_{p.} \\ \boldsymbol{\rho}_{h.} \end{pmatrix} = \begin{bmatrix} \mathbf{Q} & \mathbf{1} & \mathbf{0} \\ \mathbf{Q} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \\ \theta_h \end{pmatrix} \tag{8}$$

$$= \left[\begin{array}{ccc} \mathbf{Q}_1 & \mathbf{1} & \mathbf{0} \\ \mathbf{Q}_2 & \mathbf{1} & \mathbf{0} \\ \hline \mathbf{Q}_1 & \mathbf{0} & \mathbf{1} \\ \mathbf{Q}_2 & \mathbf{0} & \mathbf{1} \end{array}\right] \begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \\ \theta_h \end{pmatrix}, \tag{9}$$

where, in Equation 9, $\mathbf{Q}$ is partitioned as $\mathbf{Q}^T = \left[\mathbf{Q}_1^T, \mathbf{Q}_2^T\right]$ with $\mathbf{Q}_1$ corresponding to the first occasion, and $\mathbf{Q}_2$ to the second occasion.

The situation in the motivating example is different. In a non-equivalent groups design, a person does not belong to more than one group. Consequently, the log-odds for two persons p and h in different populations are:

$$\begin{pmatrix} \boldsymbol{\rho}_{p.} \\ \boldsymbol{\rho}_{h.} \end{pmatrix} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{1} & \mathbf{0} \\ \mathbf{Q}_2 & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \\ \theta_h \end{pmatrix}, \tag{10}$$

where, $\mathbf{Q}_1$ corresponds to the items administered in group 1, and $\mathbf{Q}_2$ to the items administered in group 2. Compared to Equation 9, the rows corresponding to combinations of items and occasions/populations that are not observed have been deleted.

If we partition the matrices of the two models in the motivating example, and look at the relation between:

$$\mathbf{Q}^{++} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{1} & \mathbf{0} \\ \mathbf{Q}_2 & \mathbf{0} & \mathbf{1} \end{bmatrix} \quad \text{and} \quad \mathbf{W}^{++} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{1} & \mathbf{0} \\ \mathbf{W}_2 & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

we find that these matrices do span the same column-space; i.e., the augmented matrix $[\mathbf{Q}^{++}, \mathbf{W}^{++}]$ is of rank 5. Hence, the models *are* equivalent, as we set out to prove. Specifically, there is a non-singular matrix

$$\mathbf{A} = (\mathbf{W}^{++})^{g_1}\mathbf{Q}^{++} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & 1 \end{pmatrix} \tag{11}$$

such that

$$\mathbf{W}^{++}\mathbf{A}\begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \\ \theta_h \end{pmatrix} = \mathbf{W}^{++}\begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p^* \\ \theta_h^* \end{pmatrix} = \mathbf{Q}^{++}\begin{pmatrix} -\boldsymbol{\eta} \\ \theta_p \\ \theta_h \end{pmatrix},$$

where $\theta_p^* - \theta_h^* = \theta_p - \theta_h - \tau$. The parameter τ represents differences in the relation between the abilities in the two populations. Note, that by choosing another anchor, we have in effect changed this relationship, as mentioned earlier.

7. Conclusion

We have demonstrated that the two models under consideration are equivalent, which is what we wanted to show. Intertestingly, we found that the models are equivalent in a design with non-equivalent groups while they would not be equivalent in an experimental design where the groups are samples from the same population.

The fact that it depends on the design whether or not two LLTMs are equivalent was overlooked in previous studies on equivalence of LLTMs and is of independent interest. In general, the equivalence of the two LLTMs could be investigated by looking at the relation between the parameters and the log-odds. In principle, we should consider log-odds for *all* observations:

$$\boldsymbol{\rho} = \mathbf{Q}^{++} \begin{pmatrix} -\boldsymbol{\eta} \\ \boldsymbol{\theta} \end{pmatrix}, \tag{12}$$

where $\mathbf{Q}^{++}$ depends on the design and the $\mathbf{Q}$-matrix: the design specified in terms of the virtual items. To investigate whether, given a design, two (identifiable) LLTMs are equivalent we determine whether $\mathbf{Q}^{++}$ and $\mathbf{W}^{++}$ span the same column space. If we follow the previous work, we would consider the sub-matrix $\mathbf{Q}^{+} = [\mathbf{Q}, \mathbf{1}]$. However, $\mathbf{Q}^{+}$ would only span the same column space as $\mathbf{Q}^{++}$ when the design is complete (see Eq. 8). When all persons answer to all virtual items, we only need to consider the log-odds of one person (see Eq. 4) and two models are equivalent *iff* the log-odds for one person are the same under the two models. This is not necessarily true when the design is incomplete. To prove the equivalence of the models in the motivating example, it was necessary to consider the log-odds (in Eq. 10) of at least two persons, each from a different population. We leave it as a topic for future research to develop the theory for general designs.

In closing, we mention two further topics for future research. First, depending on the design, there are combinations of items and persons for which there are no observations. When two LLTMs are equivalent, given a particular design, we may ask whether (and how) the design can be changed such that the models are no longer equivalent and it becomes possible to determine which model describes the data best. This would make an interesting and positive topic for future research. Unfortunately, for the two models in the motivating example, we would have to add persons from the first group as members of the second group, which is impossible. Second, the LLTM is a very simple model. One would expect to find similar issues in studies using other measurement models which allow more intricate forms of DIF. For example, in investigations of *partial measurement invariance* in structural equation modelling (Byrne, Shavelson, & Muthén, 1989). Unfortunately, we could not find a reference in the literature.[3] This is another topic for future research.

[3]MacCallum, Wegener, Uchino, and Fabrigar (1993) or Raykov (1997) come very close.

References

Bechger, T. M., Verstralen, H. H. F. M., & Verhelst, N. D. (2002). Equivalent linear logistic test models. *Psychometrika*, *67*, 123-136.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.

Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, *48*, 3-26.

Fischer, G. H. (2000). Gain scores revisited under an IRT perspective. In A. Boomsma, M. A. J. van Duijn, & A. B. Snijders (Eds.), *Essays on item response theory* (p. 43-68). New-York: Springer.

Fischer, G. H. (2004). Remarks on 'Equivalent linear logistic test models' by Bechger, Verstralen, and Verhelst (2002). *Psychometrika*, *69*, 305-315.

Fischer, G. H. (2007). Rasch models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (Vol. 26, p. 515-585). Amsterdam, The Netherlands: Elsevier.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185-199.

Pringle, R. M., & Rayner, A. A. (1971). *Generalized inverse matrices with applications to statistics.* London: Charles Griffin & Co.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: The Danish Institute of Educational Research. (Expanded edition, 1980. Chicago, The University of Chicago Press)

Raykov, T. (1997). Equivalent structural equation models and group equality constraints. *Multivariate Behavioral Research*, *32*(2), 95-104.

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

8. Appendix: Proof of Proposition 2

Lemma 4. *Let $r()$ denote the rank of a matrix. Then,*

$$r\left(\left[\mathbf{Q}^{+}, \mathbf{W}^{+}\right]\right) = r(\mathbf{Q}^{+}) = r(\mathbf{W}^{+}) \Leftrightarrow \mathbf{Q}^{+} \textit{ and } \mathbf{W}^{+} \textit{ span the same column space}$$

Proof. $\Leftarrow$: Assume that $\mathbf{Q}^{+}$ and $\mathbf{W}^{+}$ span the same column space. This means that each matrix is in the column space of the other. If $\mathbf{W}^{+}$ is in the column space of $\mathbf{Q}^{+}$ there exists a matrix $\mathbf{A}$ such that $\mathbf{Q}^{+} = \mathbf{W}^{+}\mathbf{A}$. Hence, the rank of $[\mathbf{Q}^{+}, \mathbf{W}^{+}]$ is the rank of $[\mathbf{W}^{+}\mathbf{A}, \mathbf{W}^{+}]$, which equals the rank of $\mathbf{W}^{+}$. The same argument with the roles of $\mathbf{Q}^{+}$ and $\mathbf{W}^{+}$ interchanged leads to the conclusion that $r([\mathbf{Q}^{+}, \mathbf{W}^{+}]) = r(\mathbf{Q}^{+}) = r(\mathbf{W}^{+})$.

$\Rightarrow$: Assume that $r([\mathbf{Q}^{+}, \mathbf{W}^{+}]) = r(\mathbf{Q}^{+}) = r(\mathbf{W}^{+})$. If $r([\mathbf{Q}^{+}, \mathbf{W}^{+}]) = r(\mathbf{Q}^{+})$, $\mathbf{W}^{+}$ is in the column space of $\mathbf{Q}^{+}$. Similarly, $r([\mathbf{Q}^{+}, \mathbf{W}^{+}]) = r(\mathbf{W}^{+})$ implies that $\mathbf{W}^{+}$ is in the column space of $\mathbf{Q}^{+}$. □

Proposition 2 now follows if the LLTMs are identified and $r(\mathbf{Q}^{+}) = r(\mathbf{W}^{+}) = p + 1$, where p denotes the number of basic parameters.

Proposition 2 is similar to Proposition 3 in Fischer (2004). Fischer's proposition consists of two parts: First, it states that two identifiable LLTMs with weight matrices $\mathbf{W}$ and $\mathbf{Q}$, both of dimension $k \times p$ are equivalent iff their weight matrices span the same column space. Second,

$$r\left(\left[\mathbf{Q}^{+}, \mathbf{W}^{+}\right]\right) = p + 1 \Leftrightarrow \mathbf{Q}^{+} \text{ and } \mathbf{W}^{+} \text{ span the same column space}$$

The first statement is a definition of equivalence and the second statement is not true. Furthermore, Fischer's proof is not a proof of Proposition 2 in this paper. Rather, he proves that $\mathbf{Q}^{+}$ and $\mathbf{W}^{+}$ span the same column space *iff* $\mathbf{Q}_a$ and $\mathbf{W}_a$ span the same column space, where $\mathbf{Q}_a$ and $\mathbf{W}_a$ are the weight matrices after normalization (see Bechger et al., 2002).