Chapter 4
Modeling Parental Involvement

In this chapter, we first outline the models used and the estimation and testing procedures employed, and then summarize the results revealed by these models.

4.1 Estimation and Testing Procedures

The procedures we used for parameter estimation and evaluation of model fit are based on marginal maximum likelihood (MML). Most of the procedures we discuss are documented in more detail elsewhere (see Bock and Aitkin 1981; Bock et al. 1988; Gibbons and Hedeker 1992; Glas 1999; Adams and Wu 2006; De Jong et al. 2007; Jennrich and Bentler 2011; Glas and Jehangir 2014). We used the public domain software package MIRT (Glas 2010) in the calculations. Additional estimation and testing procedures were used for the bi-factor model, with unidimensional models as special cases, and random item parameters as a generalization.

4.1.1 MML Estimation

The bi-factor model used in this study was in two parts: a measurement model (i.e., an IRT model) and a structural model. The measurement model pertains to a polytomously-scored response of a student n to an item i. The possible item scores range from 0 to m_i and the score of student n on item i is denoted by the variables x_{nij} ($j = 1, \ldots, m_i$) where $x_{nij} = 1$ if the response is in category 1 and zero otherwise. Note that m_i has an index i, which indicates that the maximum score of items can differ.

We describe the procedure for the bi-factor model, combined with the partial credit model (PCM; Masters 1982) and generalized partial credit model (GPCM; Muraki 1992) as IRT models, since these two models were the ones we selected for the present study. However, the theory also applies to other IRT models, such as the unidimensional PCM and GPCM, the graded response model (Samejima 1969), the

R. Annemiek Punter et al., Psychometric Framework for Modeling Parental Involvement and Reading Literacy 2016
R.A. Punter et al., *Psychometric Framework for Modeling Parental Involvement and Reading Literacy*, IEA Research for Education 1, DOI 10.1007/978-3-319-28064-6_4

sequential model (Tutz 1990), and other versions of these models with random item parameters instead of fixed item parameters.

In the bi-factor GPCM, the probability of scoring in category j ($j = 0, \ldots, m_i$) is given by

$$p_{ij}(\theta_n) = p(x_{nij} = 1|\theta_n, a, b) = \frac{\exp\left(\sum_{h=1}^{j} a_{i0}\theta_{n0} + a_{ig(n)}\theta_{ng(n)} - b_{ih}\right)}{1 + \sum_{k=1}^{m_i} \exp\left(\sum_{h=1}^{k} a_{i0}\theta_{n0} + a_{ig(n)}\theta_{ng(n)} - b_{ik}\right)} \tag{4.1}$$

where, θ_{n0} is the score of a student n on the latent scale pertaining to all countries, $\theta_{ng(n)}$ is the score on a country specific latent dimension, and the index $g(n)$ indicates the country to which student n belongs. Further, a_{i0} and $a_{ig(n)}$ are the factor loadings of item i on these two dimensions, and b_{ih} ($h = 1, \ldots, m_i$) is the item location parameter. The location parameter b_{ih} is the position on the latent scale, where it is assumed that summations such as $h = 1$ to 0 result in zero. The unidimensional GPCM lacks the country-specific dimensions $\theta_{ng(n)}$ and the associated factor loadings $a_{ig(n)}$. Further, the PCM is obtained by fixing all item parameters a_{i0} to one.

The formula for the response probability and subsequent derivations can be simplified by introducing the re-parametrization $d_{ij} = \Sigma_{h=1}^{j} b_{ih}$ and by defining $a_{ig}^t\theta_n$ as the inner product of the vectors $(a_{i0}, a_{ig(n)})$ and $(\theta_{n0}, \theta_{ng(n)})$, respectively. Thus, Eq. (4.1) becomes

$$p_{ij}(\theta_n) = \frac{\exp\left(ja_{ig}^t\theta_n - d_{ij}\right)}{1 + \sum_{k=1}^{m_i} \exp\left(ka_{ig}^t\theta_n - d_{ik}\right)} \tag{4.2}$$

The θ_0-dimension is the general dimension that pertains to all countries and is the basis for the comparison of the countries. The θ_g-dimensions are the country-specific dimensions, and the factor loadings on these dimensions give an indication of country-by-item interaction. It is assumed that within each country, the dimensions θ_0 and θ_g have a bi-variate normal distribution $N(\theta_{n0}, \theta_{ng}; \mu_g, \Sigma_g)$. For the two-dimensional country mean $\mu_g = (\mu_{g0}, \mu_g)$, it holds that the mean on the second dimension is fixed at zero, that is $\mu_g = 0$. The covariance matrix is given by

$$\Sigma_g = \begin{bmatrix} \sigma_g^2 & 0 \\ 0 & 1 \end{bmatrix}$$

In the unidimensional GPCM and PCM, the latent student parameters θ_0 have a univariate normal distribution with a mean μ_g and a variance σ_g^2. Finally, random item parameters are obtained by introducing independent multivariate normal

distributions on the parameters for each item (for further details, please consult De Jong et al. 2007).

The present application of the bi-factor model is not standard, but an extension of the basic model. Thus, the technical details on the estimation equations, expressions for the covariance matrix of the estimates, and tests of model fit, are also provided (see Appendix A).

4.1.2 *Detection and Modeling of Differential Item Functioning*

Part of the process of establishing the construct validity of a scale may consist of showing that the scale fits an IRT model. In the present study, the focus is on country-specific CDIF. CDIF can be detected using Lagrange multiplier (LM) test statistics (Rao 1947; see also, Aitchison and Silvey 1958) and CDIF can be modeled using country-specific item parameters. Glas and Jehangir (2014) already showed the feasibility of the method using PISA data, although in the slightly simpler framework of one-dimensional IRT models. The method is implemented in the public domain software package MIRT (Glas 2010). LM tests have been previously applied to IRT frameworks (Glas 1999; Glas and Falcón 2003; Glas and Dagohoy 2007). Our primary interest is not in the actual outcome of the LM test, because due to the very large sample sizes in educational surveys even the smallest model violation, that is, the smallest amount of differential item functioning (DIF), will be significant. The reason for adopting the framework of the LM test is that it clarifies the connection between the model violations, and observations and expectations used to detect DIF. Further, because it produces comprehensible and well-founded expressions for model expectations, the value of the LM test statistic can be used as measure of the effect size of DIF, and the procedure can be easily generalized to a broad class of IRT models.

To define the test and the associated residuals, we define a background variable

$$y_{nc} = \begin{cases} 1 & \text{if person } n \text{ belongs to country } c, \\ 0 & \text{if person } n \text{ does not belong to country } c. \end{cases}$$

The LM test targets the null-hypothesis of no DIF, namely the null-hypothesis where $\delta_i = 0$. The LM test statistic is computed using the MML estimates of the null-model, where δ_i is not estimated. The test is based on evaluation of the first-order derivatives of the marginal likelihood with respect to δ_i evaluated at $\delta_i = 0$ (see Glas 1999). If the first-order derivative in this point is large, the MML estimate of δ_i is far removed from zero, and the test is significant. If the first-order derivative in this point is small, the MML estimate of δ_i is probably close to zero and the test is not significant. The actual LM statistic is the squared first-order derivative divided by its estimated variance, and it has an asymptotic chi-squared

distribution with one degree of freedom. However, as already discussed, the primary interest is not so much in the test itself, but in the information it provides regarding the fit between the data and the model.

For a general definition of the approach, which also pertains to polytomously-scored items, the covariates y_{nc} (c = 1, …, C) should be defined. Special cases leading to specific DIF statistics are given later. The covariates may be separately observed person characteristics, but they may also depend on the observed response pattern, but without the response to the item i targeted.

The LM approach can be outlined using the bi-factor GPCM; the special cases for the unidimensional PCM and GPCM are obtained if the restrictions denoted above are invoked. The probability of a response is given by a generalization of the bi-factor GPCM, namely,

$$p_{ij}(\theta_n) = \frac{\exp\left(ja_{ig}^t\theta_n - d_{ij} + j\sum_c y_{nc}\delta_{ic}\right)}{1 + \sum_{k=1}^{m_i}\exp\left(ka_{ig}^t\theta_n - d_{ik} + k\sum_c y_{nc}\delta_{ic}\right)}$$

For one so-called reference country, the covariate y_{nv} is equal to zero. This country serves as a baseline where the bi-factor GPCM with item parameters a and b holds. In the other C-1 countries, the covariates y_{nv} are equal to one. It can be shown (see Glas 1999) that the test statistic is based on the residuals

$$\frac{\sum_{n=1}^{N}\sum_{j=1}^{m_i} y_{nc} jX_{ij}}{\sum_{n=1}^{N} y_{nc}} - \frac{\sum_{n=1}^{N}\sum_{j=1}^{m_i} y_{nc} jE\left(P_{ij}(\theta_n)|x_n;\lambda\right)}{\sum_{n=1}^{N} y_{nc}} \tag{4.3}$$

for c = 1, …, C-1. Dividing this residual by the number of respondents $\Sigma_n y_{nc}$ produces residuals that are the differences between the observed and expected average item-total score in country c = 1, …, C-1. The residual gauges so-called uniform DIF, in other words, the residual indicates whether the item total function (ITF) $\Sigma_j\, jP_{ij}(\theta)$ is shifted for the item, namely whether there is item-by-country interaction.

The LM statistic for the null-hypothesis $\delta_i = 0$ (c = 1, …, C-1) is a quadratic form in the (C-1)-dimensional vector of residuals and the inverse of their covariance matrix (for details, see Glas 1999). It has an asymptotic chi-squared distribution with C-1 degrees of freedom.

A special case of this procedure is obtained if one country serves as the focal country and all other countries serve as reference. Then the model under the alternative hypothesis has only one additional parameter, δ_i, and the associated LM statistic has an asymptotic chi-squared distribution with one degree of freedom.

Items that show the worst misfit, based on their value of the LM statistic and residuals, are given country-specific item parameters. From a practical point of

view, defining country-specific item parameters is equivalent to defining an incomplete design where the DIF item is split into a number of virtual items, and where each virtual item is considered as administered in a specific country. The resulting design can be analyzed using IRT software that supports the analysis of data collected in an incomplete design. We here refer to items with country-specific parameters as split items.

The method is motivated by the assumption that a substantial part of the items function the same in all countries and a limited number of items have CDIF. In the IRT model, it is assumed that all items pertain to the same latent variable θ. Items without CDIF have the same item parameters in every country. However, items with CDIF have item parameters that differ across countries. These items refer to the same latent variable θ as all the other items, but their location on the scale differs across countries. For instance, the number of cars in the family may be a good indicator of wealth, but the actual number of cars at a certain level of wealth may vary across countries, or even within countries. Having a car in the inner city of Amsterdam is clearly a sign of wealth, but, in the rural eastern part of the Netherlands, an equivalent level of wealth would probably result in the ownership of three cars.

The number of items given country-specific item parameters is a matter of choice where two considerations are relevant. First, there should remain a sufficient number of anchor items in the scale. Second, the model including the split items should fit the data. DIF statistics no longer apply to the split items. However, the fit of the item response curve of an individual item, say item *i,* can be evaluated using the test for non-uniform DIF described earlier, but using a model including country-specific items parameters. So, in this application too, test-score ranges are used as proxies for locations on the θ scale, and the test evaluates whether the model with the country-specific item parameters can properly predict the ITF.

4.2 Results of Modeling Country-Specific Differential Item Functioning

We here provide descriptive statistics at country level for each of the five parental involvement components under the PCM and GPCM, including sample size and estimated global reliability (Tables 4.1, 4.2, 4.3, 4.4 and 4.5). Sample sizes for the first four components (early literacy activities, help with homework, school practices on parental involvement from a parental perspective, and parental involvement from a student perspective) were taken from the PIRLS home and student data, providing a significantly larger sample than that available for the last component (school practices on parental involvement, school perspective), where data were derived from the PIRLS school questionnaire. The GPCM rarely improved global reliability. Components 1 (early literacy activities), 2 (help with homework), and 5 (school practices on parental involvement, school perspective) were evaluated using nine, eight, and 15 items, respectively (see also Table 3.2). Their global reliability is

Table 4.1 Country characteristics component 1: early literacy activities before beginning primary school

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Azerbaijan, Republic of	4509	6.56	0.44	1.05	0.74	0.36	0.98	0.74
Australia	3232	4.46	−0.55	1.30	0.77	−0.49	1.19	0.77
Austria	4393	5.90	0.10	1.01	0.73	0.08	0.94	0.74
Belgium (French)	3383	6.46	0.30	1.01	0.74	0.29	0.94	0.74
Bulgaria	5137	6.10	0.12	1.57	0.84	0.12	1.46	0.85
Canada	18848	4.57	−0.49	1.25	0.76	−0.44	1.14	0.76
Chinese Taipei	4242	8.41	0.98	1.11	0.78	0.90	1.03	0.78
Colombia	3798	5.79	0.11	1.19	0.77	0.13	1.10	0.77
Croatia	4539	4.62	−0.38	0.97	0.69	−0.35	0.90	0.69
Czech Republic	4397	5.28	−0.10	0.90	0.68	−0.09	0.84	0.69
Denmark	4322	6.10	0.18	0.96	0.72	0.18	0.90	0.73
Finland	4423	6.23	0.24	0.80	0.65	0.24	0.74	0.65
France	4111	5.94	0.12	1.02	0.74	0.11	0.95	0.74
Georgia	4640	4.46	−0.44	1.11	0.72	−0.44	1.02	0.72
Germany	3197	5.56	−0.01	0.96	0.71	−0.02	0.89	0.71
Hong Kong, SAR	3604	8.45	1.01	0.97	0.73	0.91	0.90	0.74
Hungary	4912	5.27	−0.11	0.92	0.69	−0.12	0.85	0.69
Indonesia	4588	6.90	0.48	1.02	0.74	0.45	0.94	0.75
Iran, Islamic Republic of	5653	7.82	0.82	1.06	0.76	0.75	0.99	0.76
Ireland	4268	4.58	−0.47	1.24	0.76	−0.43	1.14	0.76
Israel	3261	4.81	−0.33	1.11	0.74	−0.30	1.03	0.74
Italy	3873	4.97	−0.23	1.00	0.71	−0.20	0.93	0.71
Lithuania	4406	5.67	0.04	0.96	0.71	0.01	0.90	0.71
Malta	3274	5.24	−0.18	1.14	0.76	−0.17	1.06	0.76
Netherlands	2273	5.53	−0.03	0.96	0.71	−0.02	0.89	0.71
New Zealand	3357	4.37	−0.60	1.33	0.77	−0.54	1.22	0.77
Norway	2909	5.76	0.06	0.97	0.71	0.08	0.90	0.72
Northern Ireland	2107	4.02	−0.74	1.28	0.75	−0.68	1.18	0.75
Poland	4920	5.06	−0.20	0.99	0.71	−0.20	0.92	0.71
Portugal	3887	5.76	0.05	1.09	0.75	0.04	1.01	0.76
Qatar	3650	6.49	0.35	1.08	0.75	0.30	1.00	0.76
Romania	4535	5.59	−0.12	1.57	0.83	−0.12	1.46	0.84
Russian Federation	4412	4.02	−0.70	1.19	0.73	−0.68	1.09	0.73
Saudi Arabia	4369	6.52	0.36	1.04	0.75	0.36	0.97	0.75
Singapore	6194	7.16	0.51	1.24	0.80	0.47	1.15	0.81
Slovak Republic	5481	5.02	−0.24	1.08	0.74	−0.23	1.00	0.74
Slovenia	4313	4.78	−0.33	1.02	0.71	−0.31	0.94	0.71
Spain	7945	5.13	−0.18	1.03	0.72	−0.16	0.95	0.73

(continued)

Table 4.1 (continued)

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Sweden	4013	6.06	0.15	1.03	0.74	0.15	0.96	0.75
Trinidad and Tobago	3497	4.85	−0.33	1.17	0.75	−0.29	1.08	0.76
United Arab Emirates	13305	6.52	0.35	1.03	0.74	0.32	0.96	0.75

Note N is the sample size, and $\bar{X}$ the observed mean score on the component. $\mu(\theta)$ is the estimated mean, $\sigma(\theta)$ is the standard deviation, and ρ is the estimated global reliability under the partial credit model (PCM) or the generalized partial credit model (GPCM)

Table 4.2 Country characteristics component 2: help with homework

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Azerbaijan, Republic of	4541	2.99	−0.95	2.02	0.76	−0.63	1.30	0.76
Australia	3234	5.27	0.53	1.23	0.79	0.33	0.80	0.80
Austria	4430	6.26	0.83	1.22	0.81	0.57	0.81	0.82
Belgium (French)	3356	3.58	−0.44	1.74	0.78	−0.30	1.16	0.79
Bulgaria	5126	4.82	−0.22	2.28	0.83	−0.13	1.50	0.84
Canada	18844	3.99	−0.04	1.41	0.77	−0.02	0.92	0.78
Chinese Taipei	4244	5.73	0.53	1.52	0.83	0.33	1.00	0.84
Colombia	3824	3.03	−0.72	1.74	0.75	−0.46	1.12	0.76
Croatia	4532	5.08	0.44	1.28	0.79	0.32	0.88	0.82
Czech Republic	4418	4.42	0.30	1.10	0.73	0.22	0.74	0.76
Denmark	4303	5.32	0.54	1.23	0.79	0.36	0.82	0.80
Finland	4410	8.31	1.45	0.92	0.78	0.96	0.64	0.80
France	4115	3.63	−0.23	1.48	0.76	−0.15	0.99	0.78
Georgia	4622	3.05	−0.83	1.90	0.76	−0.53	1.23	0.77
Germany	3195	6.05	0.72	1.33	0.82	0.49	0.90	0.84
Hong Kong, SAR	3609	5.94	0.49	1.70	0.85	0.28	1.13	0.85
Hungary	4903	3.91	−0.26	1.71	0.79	−0.15	1.13	0.80
Indonesia	4577	3.99	−0.23	1.70	0.79	−0.21	1.10	0.79
Iran, Islamic Republic of	5650	4.68	0.16	1.53	0.80	0.07	1.01	0.81
Ireland	4268	2.99	−0.69	1.68	0.75	−0.46	1.11	0.76
Israel	3271	5.84	0.63	1.38	0.82	0.43	0.92	0.83
Italy	3867	3.78	−0.22	1.57	0.78	−0.12	1.05	0.80
Lithuania	4395	5.49	0.53	1.35	0.81	0.35	0.92	0.83
Malta	3285	4.23	0.06	1.41	0.78	0.04	0.93	0.79
Netherlands	2280	9.36	1.63	1.10	0.83	1.09	0.76	0.85
New Zealand	3351	5.28	0.43	1.44	0.81	0.26	0.94	0.82
Norway	2105	2.40	−0.82	1.41	0.69	−0.54	0.93	0.70
Northern Ireland	2908	3.56	−0.20	1.39	0.75	−0.15	0.91	0.76

(continued)

Table 4.2 (continued)

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Poland	4923	3.82	−0.10	1.40	0.76	−0.05	0.94	0.78
Portugal	3889	3.82	−0.32	1.72	0.79	−0.23	1.14	0.80
Qatar	3653	3.20	−0.54	1.63	0.76	−0.35	1.04	0.76
Romania	4533	3.71	−0.81	2.32	0.80	−0.50	1.52	0.81
Russian Federation	4417	3.39	−0.40	1.56	0.76	−0.28	1.01	0.77
Saudi Arabia	4256	3.79	−0.33	1.74	0.79	−0.23	1.12	0.79
Singapore	6190	5.83	0.56	1.51	0.83	0.33	0.99	0.84
Slovak Republic	5489	4.99	0.31	1.47	0.81	0.22	0.99	0.83
Slovenia	4340	4.78	0.25	1.42	0.80	0.18	0.96	0.82
Spain	7945	3.15	−0.67	1.76	0.76	−0.43	1.17	0.77
Sweden	3985	4.78	0.31	1.34	0.79	0.19	0.88	0.80
Trinidad and Tobago	3499	2.41	−1.09	1.76	0.72	−0.68	1.12	0.73
United Arab Emirates	13287	3.12	−0.61	1.67	0.76	−0.40	1.07	0.76

Note N is the sample size, and $\bar{X}$ the observed mean score on the component. $\mu(\theta)$ is the estimated mean, $\sigma(\theta)$ is the standard deviation, and ρ is the estimated global reliability under the partial credit model (PCM) or the generalized partial credit model (GPCM)

Table 4.3 Country characteristics component 3: school practices on parental involvement, parent perspective

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Azerbaijan, Republic of	4401	0.79	−2.02	1.40	0.47	−15.97	11.11	0.51
Australia	3185	3.51	0.39	0.13	0.04	3.12	1.04	0.36
Austria	4349	4.11	0.63	0.19	0.10	4.96	1.49	0.49
Belgium (French)	3269	4.14	0.63	0.13	0.04	0.63	1.00	0.35
Bulgaria	5029	2.70	−0.01	0.58	0.42	−0.07	4.62	0.67
Canada	18567	3.66	0.45	0.16	0.07	3.58	1.29	0.43
Chinese Taipei	4189	1.85	−0.35	0.23	0.10	−2.77	1.84	0.54
Colombia	3738	1.31	−1.31	1.31	0.55	−10.37	10.33	0.62
Croatia	4478	3.05	0.18	0.46	0.35	1.43	3.65	0.65
Czech Republic	4316	3.68	0.46	0.15	0.06	3.65	1.16	0.40
Denmark	4243	4.03	0.58	0.13	0.04	4.60	1.01	0.37
Finland	4348	4.53	0.73	0.10	0.02	5.77	0.79	0.28
France	3961	3.86	0.52	0.10	0.02	4.14	0.81	0.27
Georgia	4483	1.63	−0.80	0.96	0.51	−6.31	7.58	0.64
Germany	3097	3.88	0.54	0.13	0.04	4.25	0.99	0.36
Hong Kong, SAR	3593	1.51	−0.60	0.33	0.17	−4.75	2.64	0.59
Hungary	4793	3.38	0.36	0.22	0.13	2.83	1.77	0.52

(continued)

Table 4.3 (continued)

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Indonesia	4549	0.86	−1.65	1.09	0.42	−13.02	8.61	0.57
Iran, Islamic Republic of	5608	1.34	−1.00	0.89	0.45	−7.88	7.07	0.64
Ireland	4187	3.44	0.37	0.36	0.27	2.89	2.87	0.61
Israel	3188	2.47	−0.11	0.55	0.39	−0.87	4.34	0.64
Italy	3755	3.61	0.43	0.11	0.03	3.44	0.86	0.28
Lithuania	4347	3.45	0.38	0.22	0.13	2.99	1.77	0.51
Malta	3188	2.35	−0.25	0.80	0.51	−1.99	6.34	0.66
Netherlands	2265	4.39	0.70	0.13	0.04	5.56	1.01	0.39
New Zealand	3362	3.56	0.42	0.23	0.13	3.29	1.78	0.51
Norway	2091	3.92	0.55	0.19	0.10	4.37	1.51	0.48
Northern Ireland	2884	3.49	0.38	0.11	0.03	3.04	0.89	0.29
Poland	4790	3.25	0.32	0.15	0.05	2.50	1.15	0.38
Portugal	3745	3.60	0.43	0.11	0.03	3.44	0.86	0.28
Qatar	3610	1.87	−0.39	0.40	0.25	−3.06	3.18	0.61
Romania	4477	2.00	−0.52	0.90	0.53	−4.10	7.13	0.66
Russian Federation	4331	3.25	0.31	0.19	0.09	2.44	1.48	0.47
Saudi Arabia	4306	1.39	−1.05	1.03	0.50	−8.29	8.19	0.64
Singapore	6145	2.03	−0.23	0.18	0.07	−1.83	1.42	0.45
Slovak Republic	5344	3.08	0.23	0.23	0.13	1.83	1.79	0.52
Slovenia	4246	4.00	0.55	0.09	0.02	4.32	0.74	0.23
Spain	7699	3.53	0.41	0.20	0.11	3.27	1.62	0.49
Sweden	3974	3.59	0.43	0.13	0.04	3.36	1.02	0.35
Trinidad and Tobago	3328	2.03	−0.51	0.96	0.55	−4.07	7.57	0.67
United Arab Emirates	13061	1.68	−0.59	0.62	0.37	−4.68	4.89	0.64

Note N is the sample size, and $\bar{X}$ the observed mean score on the component. $\mu(\theta)$ is the estimated mean, $\sigma(\theta)$ is the standard deviation, and ρ is the estimated global reliability under the partial credit model (PCM) or the generalized partial credit model (GPCM)

generally >0.70, which is an acceptable level for country inferences. A value of 0.80 is generally considered an acceptable reliability level for individual inferences, and for many combinations of components and countries, this level was attained. Components 3 (school practices on parental involvement, parental perspective) and 4 (parental involvement from a student perspective), were evaluated using three items and five items, respectively; the global reliability of these estimates was thus correspondingly lower.

We also investigated the item characteristics for each component (Tables 4.6, 4.7, 4.8, 4.9 and 4.10). Local reliability, namely the extent to which different θ-values can be distinguished, was assessed using the "slope" parameter. The relatively high

Table 4.4 Country characteristics component 4: student perception of parental involvement

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Azerbaijan, Republic of	4330	1.50	−0.48	0.99	0.51	−0.54	1.18	0.51
Australia	5997	3.31	0.44	0.67	0.57	0.55	0.81	0.58
Austria	4571	1.90	−0.24	0.90	0.54	−0.29	1.09	0.55
Belgium (French)	3680	2.11	−0.12	0.87	0.55	−0.17	1.09	0.57
Bulgaria	5191	2.36	−0.24	1.18	0.64	−0.29	1.44	0.64
Canada	22750	2.46	0.08	0.79	0.56	0.13	0.96	0.57
Chinese Taipei	4276	4.36	0.65	0.84	0.69	0.79	1.04	0.70
Colombia	3793	1.42	−0.74	1.17	0.53	−0.88	1.40	0.53
Croatia	4564	2.08	−0.04	0.74	0.50	−0.06	0.88	0.50
Czech Republic	4483	1.38	−0.52	0.87	0.45	−0.62	1.06	0.47
Denmark	4543	2.58	0.21	0.66	0.51	0.25	0.80	0.52
England	3912	3.30	0.47	0.61	0.53	0.54	0.73	0.54
Finland	4599	3.57	0.55	0.58	0.53	0.67	0.70	0.55
France	4403	2.31	0.01	0.80	0.55	−0.02	1.01	0.57
Georgia	4581	1.56	−0.53	1.05	0.53	−0.62	1.25	0.53
Germany	3600	1.90	−0.16	0.80	0.50	−0.13	0.97	0.53
Hong Kong, SAR	3826	5.34	0.93	0.70	0.67	1.10	0.88	0.68
Hungary	5105	1.95	−0.23	0.91	0.54	−0.31	1.09	0.54
Indonesia	4662	2.32	−0.04	0.90	0.58	−0.01	1.10	0.60
Iran, Islamic Republic of	5727	2.14	−0.07	0.83	0.54	−0.08	0.98	0.54
Ireland	4415	2.27	0.00	0.80	0.54	0.03	0.97	0.56
Israel	4117	2.46	0.06	0.83	0.57	0.08	1.00	0.58
Italy	4100	2.17	−0.01	0.76	0.52	−0.07	0.96	0.54
Lithuania	4591	2.17	−0.02	0.77	0.52	−0.04	0.93	0.53
Malta	3519	2.53	0.09	0.81	0.57	0.14	0.98	0.59
Netherlands	3955	3.56	0.48	0.72	0.61	0.52	0.89	0.61
New Zealand	5549	3.03	0.36	0.67	0.55	0.44	0.80	0.56
Northern Ireland	3523	2.37	0.16	0.61	0.46	0.17	0.75	0.47
Norway	3112	2.50	0.19	0.65	0.49	0.29	0.80	0.53
Poland	4953	2.20	−0.05	0.82	0.54	−0.10	1.01	0.55
Portugal	4037	1.91	−0.19	0.83	0.52	−0.26	1.02	0.53
Qatar	3947	2.82	0.17	0.89	0.62	0.19	1.08	0.62
Romania	4592	1.69	−0.57	1.16	0.57	−0.71	1.39	0.56
Russian Federation	4444	1.82	−0.25	0.86	0.51	−0.31	1.05	0.53
Saudi Arabia	4425	2.55	0.06	0.90	0.60	0.07	1.08	0.61
Singapore	6275	4.25	0.66	0.74	0.65	0.77	0.92	0.66
Slovak Republic	5586	1.76	−0.45	1.06	0.56	−0.57	1.29	0.56
Slovenia	4456	2.13	−0.02	0.75	0.51	−0.02	0.91	0.52
Spain	8501	2.07	−0.15	0.88	0.55	−0.18	1.07	0.56

(continued)

Table 4.4 (continued)

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Sweden	4533	2.45	0.18	0.64	0.49	0.23	0.79	0.51
Trinidad and Tobago	3875	1.52	−0.65	1.12	0.54	−0.75	1.34	0.55
United Arab Emirates	14209	2.23	−0.11	0.94	0.58	−0.14	1.14	0.59
United States	12501	2.72	0.15	0.84	0.60	0.20	1.02	0.61

Note N is the sample size, and $\bar{X}$ the observed mean score on the component. $\mu(\theta)$ is the estimated mean, $\sigma(\theta)$ is the standard deviation, and ρ is the estimated global reliability under the partial credit model (PCM) or the generalized partial credit model (GPCM)

Table 4.5 Country characteristics component 5: school practices on parental involvement, school perspective

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Azerbaijan, Republic of	169	32.89	0.32	0.55	0.70	0.59	0.90	0.73
Australia	269	35.41	0.64	0.64	0.74	1.04	1.06	0.76
Austria	158	31.00	0.05	0.47	0.64	−0.14	0.81	0.71
Belgium (French)	118	23.37	−0.80	0.59	0.73	−1.13	0.92	0.74
Bulgaria	147	30.14	−0.04	0.70	0.79	0.10	1.10	0.81
Canada	1084	33.37	0.38	0.67	0.76	0.56	1.09	0.79
Chinese Taipei	150	34.35	0.54	0.88	0.83	0.77	1.48	0.85
Colombia	149	32.73	0.35	0.77	0.81	0.87	1.29	0.81
Croatia	152	30.50	−0.02	0.46	0.63	0.01	0.74	0.68
Czech Republic	174	28.28	−0.29	0.46	0.64	−0.30	0.81	0.72
Denmark	231	25.93	−0.54	0.42	0.60	−1.03	0.76	0.67
England	120	32.83	0.29	0.57	0.71	0.30	0.89	0.74
Finland	139	25.60	−0.59	0.50	0.68	−1.01	0.85	0.72
France	167	27.52	−0.35	0.57	0.73	−0.60	0.99	0.78
Georgia	171	30.85	0.07	0.69	0.79	0.24	1.17	0.82
Germany	187	30.16	−0.05	0.50	0.68	−0.19	0.82	0.72
Hong Kong, SAR	125	30.18	−0.04	0.63	0.76	−0.31	1.05	0.80
Hungary	143	29.06	−0.18	0.52	0.69	−0.14	0.85	0.73
Indonesia	155	27.53	−0.37	0.75	0.82	−0.60	1.21	0.84
Iran, Islamic Republic of	244	32.60	0.30	0.88	0.85	0.65	1.46	0.85
Ireland	145	27.75	−0.32	0.62	0.76	−0.75	1.07	0.81
Israel	132	32.24	0.25	0.71	0.79	0.31	1.12	0.81
Italy	200	27.80	−0.26	0.65	0.77	−0.38	1.08	0.80
Lithuania	151	30.42	−0.02	0.49	0.67	−0.07	0.83	0.72
Malta	93	30.99	0.09	0.59	0.73	0.02	0.89	0.75
Netherlands	117	26.97	−0.42	0.43	0.61	−0.80	0.77	0.69
New Zealand	175	34.13	0.47	0.57	0.70	0.56	0.88	0.73

(continued)

Table 4.5 (continued)

Country	N	$\bar{X}$	PCM			GPCM		
			$\mu(\theta)$	$\sigma(\theta)$	ρ	$\mu(\theta)$	$\sigma(\theta)$	ρ
Northern Ireland	117	29.23	−0.17	0.55	0.71	−0.53	0.91	0.76
Norway	115	26.03	−0.54	0.39	0.56	−0.93	0.66	0.62
Poland	148	31.57	0.15	0.53	0.69	0.22	0.83	0.71
Portugal	147	29.62	−0.10	0.62	0.76	−0.16	0.98	0.78
Qatar	166	33.86	0.53	0.96	0.85	1.02	1.59	0.85
Romania	147	32.91	0.34	0.71	0.78	0.76	1.21	0.81
Russian Federation	202	34.42	0.46	0.46	0.62	0.69	0.82	0.70
Saudi Arabia	169	26.57	−0.48	0.83	0.85	−0.55	1.43	0.88
Singapore	176	32.40	0.25	0.60	0.73	0.31	1.05	0.79
Slovak Republic	194	28.70	−0.22	0.56	0.72	−0.20	0.97	0.77
Slovenia	191	29.25	−0.14	0.47	0.65	−0.21	0.81	0.71
Spain	302	28.96	−0.18	0.65	0.78	−0.30	1.08	0.81
Sweden	132	27.33	−0.38	0.55	0.71	−0.57	0.89	0.75
Trinidad and Tobago	147	30.84	0.07	0.78	0.82	0.43	1.34	0.85
United Arab Emirates	419	32.42	0.29	0.80	0.82	0.45	1.25	0.83
United States	331	35.36	0.66	0.73	0.78	1.01	1.29	0.81

Note N is the sample size, and $\bar{X}$ the observed mean score on the component. $\mu(\theta)$ is the estimated mean, $\sigma(\theta)$ is the standard deviation, and ρ is the estimated global reliability under the partial credit model (PCM) or the generalized partial credit model (GPCM)

Table 4.6 Response frequencies and item parameter estimates under the generalized partial credit model for items in component 1: early literacy activities

Item	Slope	Intercept	$I(0)$	Relative frequency response categories		
				Cat0	Cat1	Cat2
ASBH02A	1.26	1.84	0.44	0.54	0.41	0.05
ASBH02B	1.24	1.47	0.46	0.48	0.46	0.07
ASBH02C	0.77	0.98	0.23	0.49	0.41	0.11
ASBH02D	1.09	0.85	0.45	0.43	0.44	0.14
ASBH02E	0.95	1.80	0.24	0.62	0.34	0.04
ASBH02F	1.18	0.82	0.45	0.36	0.52	0.12
ASBH02G	1.24	0.57	0.52	0.33	0.51	0.16
ASBH02H	1.06	1.12	0.38	0.47	0.44	0.10
ASBH02I	1.07	0.89	0.43	0.44	0.42	0.13

Note The latent distributions of the countries are normed to an overall mean of zero. Slope and intercept are the parameters ai0 and the mean of the location parameters bi1, bi2, …, bih, etc., respectively. $I(0)$ is the information value of the item at $\theta = 0$. Cat0, Cat1, Cat2 indicate the frequency with which item categories 0, 1 and 2 are endorsed, respectively. The components, items and corresponding category labels are described in Table 3.2

Table 4.7 Response frequencies and item parameter estimates under the generalized partial credit model for items in component 2: help with homework

Item	Slope	Intercept	$I(0)$	Relative frequency response categories			
				Cat0	Cat1	Cat2	Cat3
ASBH09A	1.17	2.44	0.21	0.78	0.18	0.03	0.02
ASBH09B	1.63	2.01	0.78	0.56	0.32	0.07	0.05
ASBH09C	1.15	2.15	0.18	0.81	0.14	0.03	0.03
ASBH09D	1.10	2.27	0.24	0.73	0.22	0.03	0.02
ASBH09E	1.56	2.51	0.31	0.77	0.17	0.03	0.04
ASBH09F	1.69	1.09	1.25	0.43	0.33	0.10	0.14
ASBH09G	2.26	1.87	1.62	0.43	0.37	0.12	0.08
ASBH09H	1.45	1.66	0.77	0.47	0.38	0.11	0.04

Note The latent distributions of the countries are normed to an overall mean of zero. Slope and intercept are the parameters ai0 and the mean of the location parameters bi1, bi2, …, bih, etc., respectively. $I(0)$ is the information value of the item at $\theta = 0$. Cat0, Cat1, Cat2, and Cat3 indicate the frequency with which item categories 0, 1, 2 and 3 are endorsed, respectively. The content of the components, items and corresponding category labels are described in Table 3.2

Table 4.8 Response frequencies and item parameter estimates under the generalized partial credit model for items in component 3: school practices on parental involvement, parent perspective

Item	Slope	Intercept	$I(0)$	Relative frequency response categories			
				Cat0	Cat1	Cat2	Cat3
ASBH10A	0.61	1.41	0.15	0.54	0.37	0.07	0.02
ASBH10B	0.61	0.36	0.34	0.30	0.31	0.23	0.16
ASBH10E	0.58	0.52	0.30	0.38	0.29	0.19	0.14

Note The latent distributions of the countries are normed to an overall mean of zero. Slope and intercept are the parameters ai0 and the mean of the location parameters bi1, bi2, …, bih, etc., respectively. $I(0)$ is the information value of the item at $\theta = 0$. Cat0, Cat1, Cat2, and Cat3 indicate the frequency with which item categories 0, 1, 2 and 3 are endorsed, respectively. The components, items and corresponding category labels are described in Table 3.2

value for PIRLS item ASBH02A (“read books”), indicates that this item of the scale performed best in this respect. Local reliability is further supported if the item location parameters agree closely with the mean of a latent distribution. In this respect, item ASBH02G (“play word games”) performed best, because the latent distributions of the countries were normed to an overall mean of zero. Together the intercept and slope parameters determine the information value of an item. Higher values for the information value of an item at $\theta = 0$, namely $I(0)$, indicate the item made a higher contribution to the local reliability of the component.

For component 1 (early literacy activities), the item ASBH02C (“sing songs”) has a lower information value than the other items. This should be taken into

Table 4.9 Response frequencies and item parameter estimates under the generalized partial credit model for items in component 4: student perception of parental involvement

Item	Slope	Intercept	$I(0)$	Relative frequency response categories			
				Cat0	Cat1	Cat2	Cat3
ASBG07A	1.01	1.47	0.32	0.67	0.21	0.05	0.07
ASBG07B	0.96	1.15	0.43	0.56	0.27	0.08	0.09
ASBG07C	0.85	1.35	0.22	0.75	0.14	0.04	0.08
ASBG07D	0.77	1.21	0.22	0.73	0.14	0.04	0.09
ASBR09C	0.55	1.55	0.09	0.76	0.18	0.04	0.02

Note The latent distributions of the countries are normed to an overall mean of zero. Slope and intercept are the parameters ai0 and the mean of the location parameters bi1, bi2, …, bih, etc., respectively. $I(0)$ is the information value of the item at $\theta = 0$. Cat0, Cat1, Cat2, and Cat3 indicate the frequency with which item categories 0, 1, 2 and 3 are endorsed, respectively. The components, items and corresponding category labels are described in Table 3.2

Table 4.10 Response frequencies and item parameter estimates under the generalized partial credit model for items in component 5: school practices on parental involvement, school perspective

Item	Slope	Intercept	$I(0)$	Relative frequency response categories				
				Cat0	Cat1	Cat2	Cat3	Cat4
ACBG11AA	0.75	−2.88	0.14	0.00	0.01	0.37	0.62	–
ACBG11AB	0.91	−2.95	0.20	0.00	0.02	0.33	0.65	–
ACBG11AC	0.87	−2.34	0.23	0.00	0.05	0.39	0.56	–
ACBG11AD	0.57	−1.34	0.14	0.03	0.07	0.29	0.62	–
ACBG11BA	0.47	−0.66	0.16	0.07	0.16	0.37	0.40	–
ACBG11BB	0.51	−0.64	0.20	0.06	0.30	0.32	0.32	–
ACBG11CA	0.70	−0.55	0.29	0.05	0.33	0.38	0.23	–
ACBG11CB	0.84	−1.29	0.38	0.03	0.18	0.35	0.44	–
ACBG11CC	1.27	−1.27	0.72	0.01	0.38	0.37	0.24	–
ACBG11CD	1.13	−1.42	0.66	0.01	0.45	0.29	0.25	–
ACBG11CE	1.09	−1.10	0.60	0.03	0.30	0.39	0.29	–
ACBG11CF	0.41	0.02	0.18	0.23	0.25	0.27	0.25	–
ACBG11CG	0.52	0.26	0.23	0.24	0.31	0.30	0.15	–
ACBG12E	0.25	−0.35	0.05	0.02	0.13	0.46	0.31	0.09
ACBG12F	0.20	−0.18	0.03	0.04	0.17	0.46	0.26	0.08

Note The latent distributions of the countries are normed to an overall mean of zero. Slope and intercept are the parameters ai0 and the mean of the location parameters bi1, bi2, …, bih, etc., respectively. $I(0)$ is the information value of the item at $\theta = 0$. Cat0, Cat1, Cat2, Cat3, and Cat4 indicate the frequency with which item categories 0, 1, 2, 3 and 4 are endorsed, respectively. The components, items and corresponding category labels are described in Table 3.2

account when redesigning the instrument for future surveys; in other words, this item may be the first candidate for replacement. Compared to component 1 (early literacy activities), the items in component 2 (helping with homework) were more informative, while items in component 3 (school practices on parental involvement, parent perspective) performed poorly. Components 4 (school practices for parental involvement from a student perspective) and 5 (school practices for parental involvement from a school perspective) provided differing results; in particular, the last two items of component 5 ("parental support for student achievement within school" and "parental involvement in school activities") performed particularly poorly.

Comparing the parameter estimates in the GPCM and the GPCM with random item parameters (henceforth the random GPCM) revealed that the agreement between the slopes and intercepts under the GPCM and the means of the slopes and intercepts under the random GPCM was high (Tables 4.11, 4.12, 4.13, 4.14 and 4.15). A higher variance provides an initial indication that the item functions differently in different countries, a topic we address in more detail later. Here, the effects are global over countries and thus only permit global inferences. For instance, for component 1, the last item, ASBH02I ("read aloud signs and tables") has the lowest CDIF because the variance of the intercepts and slopes across the countries is the lowest among the items (Table 4.11). A low variance indicates that the item parameters do not vary much across countries. Evaluating the relative CDIF of the other eight items is more difficult, because of the trade-off between the standard deviation for the slope and the intercept.

Table 4.11 Item parameter estimates under the generalized partial credit model (GPCM) and GPCM with random item parameters for items in component 1: early literacy activities

Item	GPCM		GPCM random item parameters			
	Slope	Intercept	Slope	SD (Slope)	Intercept	SD (Intercept)
ASBH02A	1.26	1.84	1.37	0.22	2.06	0.66
ASBH02B	1.24	1.47	1.25	0.15	1.50	0.31
ASBH02C	0.77	0.98	0.80	0.12	1.03	0.34
ASBH02D	1.09	0.85	1.21	0.18	0.86	0.44
ASBH02E	0.95	1.80	1.01	0.19	2.01	0.68
ASBH02F	1.18	0.82	1.33	0.23	0.93	0.42
ASBH02G	1.24	0.57	1.35	0.15	0.60	0.27
ASBH02H	1.06	1.12	1.16	0.16	1.17	0.41
ASBH02I	1.07	0.89	1.09	0.11	0.87	0.22

Note The latent distributions of the countries are normed to an overall mean of zero. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.12 Item parameter estimates under the generalized partial credit model (GPCM) and GPCM with random item parameters for items in component 2: help with homework

Item	GPCM		GPCM random item parameters			
	Slope	Intercept	Slope	SD (Slope)	Intercept	SD (Intercept)
ASBH09A	1.17	2.44	1.331	0.619	3.686	1.547
ASBH09B	1.63	2.01	1.313	0.534	2.947	1.880
ASBH09C	1.15	2.15	1.396	0.554	2.199	1.203
ASBH09D	1.10	2.27	1.227	0.314	3.736	1.610
ASBH09E	1.56	2.51	1.437	0.634	3.446	1.208
ASBH09F	1.69	1.09	1.477	0.503	0.707	1.251
ASBH09G	2.26	1.87	1.308	0.434	0.796	1.154
ASBH09H	1.45	1.66	1.559	0.224	1.518	1.210

Note The latent distributions of the countries are normed to an overall mean of zero. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.13 Item parameter estimates under the generalized partial credit model (GPCM) and GPCM with random item parameters for items in component 3: school practices on parental involvement, parent perspective

Item	GPCM		GPCM random item parameters			
	Slope	Intercept	Slope	SD (Slope)	Intercept	SD (Intercept)
ASBH10A	0.61	1.41	1.218	1.388	4.477	4.172
ASBH10B	0.61	0.36	4.144	1.601	2.751	4.923
ASBH10E	0.58	0.52	3.843	1.791	3.469	5.232

Note The latent distributions of the countries are normed to an overall mean of zero. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.14 Item parameter estimates under the generalized partial credit model (GPCM) and GPCM with random item parameters for items in component 4: student perception of parental involvement

Item	GPCM		GPCM random item parameters			
	Slope	Intercept	Slope	SD (Slope)	Intercept	SD (Intercept)
ASBG07A	1.01	1.47	0.924	0.161	1.473	1.102
ASBG07B	0.96	1.15	0.994	0.357	1.155	0.943
ASBG07C	0.85	1.35	0.989	0.316	1.937	2.614
ASBG07D	0.77	1.21	0.990	0.240	1.917	3.017
ASBR09C	0.55	1.55	0.553	0.050	2.100	2.782

Note The latent distributions of the countries are normed to an overall mean of zero. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.15 Item parameter estimates under the generalized partial credit model (GPCM) and GPCM with random item parameters for items in component 5: school practices on parental involvement, school perspective

Item	GPCM		GPCM random item parameters			
	Slope	Intercept	Slope	SD (Slope)	Intercept	SD (Intercept)
ACBG11AA	0.75	−2.88	0.689	0.664	−1.667	1.396
ACBG11AB	0.91	−2.95	1.029	0.377	−2.122	0.797
ACBG11AC	0.87	−2.34	0.998	0.506	−2.110	0.778
ACBG11AD	0.57	−1.34	0.466	1.042	−1.480	0.461
ACBG11BA	0.47	−0.66	0.645	0.876	−0.581	1.033
ACBG11BB	0.51	−0.64	0.627	0.807	−0.583	0.462
ACBG11CA	0.70	−0.55	0.887	0.491	−0.576	0.434
ACBG11CB	0.84	−1.29	0.890	0.621	−1.120	0.614
ACBG11CC	1.27	−1.27	1.236	0.620	−0.995	0.682
ACBG11CD	1.13	−1.42	1.194	0.515	−1.122	0.625
ACBG11CE	1.09	−1.10	1.132	0.229	−1.023	0.168
ACBG11CF	0.41	0.02	0.548	0.738	0.029	0.342
ACBG11CG	0.52	0.26	0.737	0.514	0.071	0.781
ACBG12E	0.25	−0.35	0.123	1.453	0.551	1.954
ACBG12F	0.20	−0.18	0.279	1.431	−0.030	1.789

Note The latent distributions of the countries are normed to an overall mean of zero. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

This pattern is repeated for component 2; the items ASBH09F ("helping child practice reading") and ASBH09G ("helping child practice math skills") performed slightly better than the other items (Table 4.12). Conversely, component 3 showed a substantial difference between the item parameters estimated with the GPCM and those estimated using the random GPCM (Table 4.13), indicating this short scale was quite unstable.

The analyses of components 4 and 5 indicated all the items performed comparably with respect to CDIF (Tables 4.14 and 4.15), although questions surrounding specific item-by-country interaction and the influence of the inferences on country means and latent regression remain unanswered.

We compared CDIF as identified by the random GPCM with CDIF as identified using the latent residuals defined by Eq. (4.3) and aggregated over countries (Tables 4.16, 4.17, 4.18, 4.19 and 4.20). Overall the agreement between the methods was high. For instance, item ASBH02I performed strongly in all methods, as did item ASBH02G (Table 4.16). In general, the residuals with the GPCM are smaller than those with the PCM, because the latter model has fewer parameters. Other studies (see e.g., Glas and Jehangir 2014) confirm this expectation. However, we found that differences between the PCM and the GPCM were very small. We

Table 4.16 Absolute differential item functioning (DIF) under the partial credit model (PCM) and the generalized partial credit model (GPCM) and standard deviation random item parameters on items in component 1: early literacy activities

Item	PCM	GPCM	SD (Slope)	SD (Intercept)
ASBH02A	0.12	0.11	0.228	0.667
ASBH02B	0.08	0.08	0.158	0.318
ASBH02C	0.09	0.10	0.126	0.349
ASBH02D	0.12	0.12	0.183	0.443
ASBH02E	0.10	0.10	0.192	0.688
ASBH02F	0.09	0.09	0.239	0.421
ASBH02G	0.07	0.07	0.155	0.279
ASBH02H	0.10	0.10	0.161	0.416
ASBH02I	0.07	0.07	0.112	0.229

Note The columns labeled PCM and GPCM give the mean residuals as estimated under the unidimensional versions of these two models. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.17 Absolute differential item functioning (DIF) under the partial credit model (PCM) and the generalized partial credit model (GPCM) and standard deviation random item parameters on items in component 2: help with homework

Item	PCM	GPCM	SD (Slope)	SD (Intercept)
ASBH09A	0.11	0.12	0.619	1.547
ASBH09B	0.07	0.07	0.534	1.880
ASBH09C	0.10	0.10	0.554	1.203
ASBH09D	0.10	0.10	0.314	1.610
ASBH09E	0.08	0.08	0.634	1.208
ASBH09F	0.14	0.12	0.503	1.251
ASBH09G	0.08	0.06	0.434	1.154
ASBH09H	0.07	0.07	0.224	1.210

Note The columns labeled PCM and GPCM give the mean residuals as estimated under the unidimensional versions of these two models. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.18 Absolute differential item functioning (DIF) under the partial credit model (PCM) and the generalized partial credit model (GPCM) and standard deviation random item parameters on items in component 3: school practices on parental involvement, parent perspective

Item	PCM	GPCM	SD (Slope)	SD (Intercept)
ASBH10A	0.13	0.47	1.388	4.172
ASBH10B	0.07	0.36	1.601	4.923
ASBH10E	0.09	0.38	1.791	5.232

Note The columns labeled PCM and GPCM give the mean residuals as estimated under the unidimensional versions of these two models. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.19 Absolute differential item functioning (DIF) under the partial credit model (PCM) and the generalized partial credit model (GPCM) and standard deviation random item parameters on items in component 4: student perception of parental involvement

Item	PCM	GPCM	SD (Slope)	SD (Intercept)
ASBG07A	0.08	0.07	0.161	1.102
ASBG07B	0.09	0.08	0.357	0.943
ASBG07C	0.07	0.08	0.316	2.614
ASBG07D	0.12	0.12	0.240	3.017
ASBR09C	0.07	0.08	0.050	2.782

Note The columns labeled PCM and GPCM give the mean residuals as estimated under the unidimensional versions of these two models. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

Table 4.20 Absolute differential item functioning (DIF) under the partial credit model (PCM) and the generalized partial credit model (GPCM) and standard deviation random item parameters on items in component 5: school practices on parental involvement, school perspective

Item	PCM	GPCM	SD (Slope)	SD (Intercept)
ACBG11AA	0.23	0.21	0.664	1.396
ACBG11AB	0.19	0.17	0.377	0.797
ACBG11AC	0.17	0.16	0.506	0.778
ACBG11AD	0.16	0.16	1.042	0.461
ACBG11BA	0.32	0.35	0.876	1.033
ACBG11BB	0.24	0.24	0.807	0.462
ACBG11CA	0.20	0.18	0.491	0.434
ACBG11CB	0.22	0.23	0.621	0.614
ACBG11CC	0.15	0.13	0.620	0.682
ACBG11CD	0.21	0.17	0.515	0.625
ACBG11CE	0.11	0.11	0.229	0.168
ACBG11CF	0.29	0.32	0.738	0.342
ACBG11CG	0.32	0.34	0.514	0.781
ACBG12E	0.26	0.27	1.453	1.954
ACBG12F	0.25	0.24	1.431	1.789

Note The columns labeled PCM and GPCM give the mean residuals as estimated under the unidimensional versions of these two models. SD (Slope) indicates the standard deviation of the slope. SD (Intercept) indicates the standard deviation of the intercept. Item descriptions are provided in Table 3.2

tentatively conclude the PCM fits the data quite well. A striking exception, again, was component 3. Here the fit of the GPCM was worse than the fit of the PCM, which leads to the conclusion that the slopes are very hard to estimate. This is in agreement with the reported low global reliability. Obviously, variance in the θ-distribution is too small to support a proper estimate of the slope parameters.

Table 4.21 Residual analysis for country-by-item interactions for component 1: early literacy activities

Country	Item									10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5	6	7	8	9			
Azerbaijan, Republic of			++		++	−−		−−	+	4	5	0.146
Australia	–									0	1	0.072
Austria	–			+						0	2	0.096
Belgium (French)										0	0	0.062
Bulgaria	+									0	1	0.058
Canada										0	0	0.059
Chinese Taipei							−−	++		2	2	0.106
Colombia	++		–						−−	2	3	0.095
Croatia	+									0	1	0.060
Czech Republic								++		1	1	0.104
Denmark	−−			++	−−					3	3	0.150
Finland	−−	+		++	−−		++			4	5	0.160
France						++				1	1	0.080
Georgia					++					1	1	0.080
Germany	–			+	–			+		0	4	0.108
Hong Kong, SAR										0	0	0.068
Hungary		−−					−−	++	++	4	4	0.148
Indonesia		++			+	–	–	−−	+	2	6	0.186
Iran, Islamic Republic of										0	0	0.067
Ireland										0	0	0.073
Israel										0	0	0.066
Italy	+									0	1	0.055
Lithuania			++							1	1	0.072

(continued)

Table 4.21 (continued)

Country	Item									10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5	6	7	8	9			
Malta									+	0	1	0.060
Netherlands	−−		–	+				+		1	4	0.129
New Zealand										0	0	0.068
Northern Ireland	–		−−	++	−−		++			4	5	0.168
Norway								+		0	1	0.086
Poland										0	0	0.040
Portugal						+				0	1	0.070
Qatar	+		++	–						1	3	0.111
Romania	++									1	1	0.074
Russian Federation			++							1	1	0.097
Saudi Arabia	++			–				–		1	3	0.129
Singapore				–	+					0	2	0.075
Slovak Republic										0	0	0.071
Slovenia				+						0	1	0.057
Spain	+					++				1	2	0.091
Sweden	–			++	−−		+			2	4	0.118
Trinidad and Tobago										0	0	0.074
United Arab Emirates	+			–						0	2	0.092

Note + indicates that residual belongs to the 20 % most positive residuals, ++ indicates that residual even belongs to the 10 % most positive residuals. − indicates that residual belongs to the 20 % most negative residuals, −− indicates that residual even belongs to the 10 % most negative residuals. The 10 % cultural differential item functioning (CDIF) and 20 % CDIF columns give the number of outliers in the two respective regions. Absolute residual refers to the means over items of the absolute values of the residuals. The content of items is described in Table 3.2

Table 4.22 Residual analysis for country-by-item interactions for component 2: help with homework

Country	Item								10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5	6	7	8			
Azerbaijan, Republic of	+								0	1	0.084
Australia			+			− −			1	2	0.097
Austria				–					0	1	0.105
Belgium (French)									0	0	0.056
Bulgaria									0	0	0.056
Canada									0	0	0.027
Chinese Taipei	++			+					1	2	0.100
Colombia									0	0	0.032
Croatia	− −		–	− −		++			3	4	0.182
Czech Republic	–				–	++		++	2	4	0.148
Denmark	–				+				0	2	0.057
Finland	–		− −	++		++			3	4	0.131
France		–						+	0	2	0.097
Georgia		+ +							1	1	0.108
Germany									0	0	0.065
Hong Kong, SAR	+			++		–	–		1	4	0.132
Hungary				–	+				0	2	0.083
Indonesia	++	+	++			− −	− −		4	5	0.188
Iran, Islamic Republic of			++					–	1	2	0.125
Ireland									0	0	0.059
Israel					+				0	1	0.076
Italy	–						+		0	2	0.098

(continued)

Table 4.22 (continued)

Country	Item								10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5	6	7	8			
Lithuania	–	–			+	++			1	4	0.134
Malta		+ +							1	1	0.083
Netherlands	++		++	– –	+ +	–	++	– –	6	7	0.249
New Zealand			++						1	1	0.097
Northern Ireland						– –			1	1	0.074
Norway									0	0	0.055
Poland	–					+			0	2	0.080
Portugal	++								1	1	0.069
Qatar									0	0	0.045
Romania									0	0	0.044
Russian Federation				++					1	1	0.068
Saudi Arabia						– –			1	1	0.089
Singapore	++			+			–	+	1	4	0.129
Slovak Republic	– –					++			2	2	0.115
Slovenia			–						0	1	0.080
Spain									0	0	0.064
Sweden									0	0	0.057
Trinidad and Tobago									0	0	0.038
United Arab Emirates									0	0	0.037

Note + indicates that residual belongs to the 20 % most positive residuals, ++ indicates that residual even belongs to the 10 % most positive residuals. − indicates that residual belongs to the 20 % most negative residuals, −− indicates that residual belongs to the 10 % most negative residuals. The 10 % cultural differential item functioning (CDIF) and 20 % CDIF columns give the number of outliers in the two respective regions. Absolute residual refers to the means over items of the absolute values of the residuals. The content of items is described in Table 3.2

Table 4.23 Residual analysis for country-by-item interactions for component 3: school practices on parental involvement, parent perspective

Country	Item			10 % CDIF	20 % CDIF	Absolute residual
	1	2	3			
Azerbaijan, Republic of			+	0	1	0.084
Australia				0	0	0.032
Austria	++			1	1	0.102
Belgium (French)	+			0	1	0.088
Bulgaria	+			0	1	0.110
Canada				0	0	0.058
Chinese Taipei				0	0	0.057
Colombia	−−			1	1	0.112
Croatia	−−			1	1	0.090
Czech Republic	++			1	1	0.085
Denmark	−−			1	1	0.071
Finland	++			1	1	0.096
France	+			0	1	0.081
Georgia	−−			1	1	0.088
Germany	++			1	1	0.164
Hong Kong, SAR	+			0	1	0.054
Hungary				0	0	0.026
Indonesia	–	+		0	2	0.142
Iran, Islamic Republic of				0	0	0.034
Ireland				0	0	0.073
Israel				0	0	0.042
Italy	+			0	1	0.106
Lithuania				0	0	0.029
Malta	−−			1	1	0.082
Netherlands				0	0	0.039
New Zealand				0	0	0.037
Northern Ireland				0	0	0.030
Norway	−−			0	0	0.104
Poland				0	0	0.050
Portugal				0	0	0.037
Qatar	+			0	1	0.075
Romania	−−			1	1	0.127
Russian Federation		+		0	1	0.088
Saudi Arabia	–			0	1	0.048
Singapore	++			1	1	0.083
Slovak Republic				0	0	0.049
Slovenia	++			1	1	0.072
Spain				0	0	0.018

(continued)

Table 4.23 (continued)

Country	Item			10 % CDIF	20 % CDIF	Absolute residual
	1	2	3			
Sweden				0	0	0.016
Trinidad and Tobago	–			0	1	0.109
United Arab Emirates				0	0	0.044

Note + indicates that residual belongs to the 20 % most positive residuals, ++ indicates that residual even belongs to the 10 % most positive residuals. − indicates that residual belongs to the 20 % most negative residuals, −− indicates that residual even belongs to the 10 % most negative residuals. The 10 % cultural differential item functioning (CDIF) and 20 % CDIF columns give the number of outliers in the two respective regions. Absolute residual refers to the means over items of the absolute values of the residuals. Item descriptions are provided in Table 3.2

Table 4.24 Residual analysis for country-by-item interactions for component 4: student perception of parental involvement

Country	Item					10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5			
Azerbaijan, Republic of						0	0	0.040
Australia						0	0	0.060
Austria						0	0	0.037
Belgium (French)				–		0	1	0.076
Bulgaria				+		0	1	0.075
Canada						0	0	0.068
Chinese Taipei	++			−−		2	2	0.117
Colombia						0	0	0.034
Croatia	–			++		1	2	0.094
Czech Republic						0	0	0.051
Denmark						0	0	0.056
England	+		–			0	2	0.088
Finland					++	1	1	0.103
France						0	0	0.068
Georgia			+			0	1	0.075
Germany	+			−−		1	2	0.146
Hong Kong, SAR		–				0	1	0.087
Hungary	–			++		1	2	0.110
Indonesia					−−	1	1	0.080
Iran, Islamic Republic of			+			0	1	0.071
Ireland		+		−−		1	2	0.112

(continued)

Table 4.24 (continued)

Country	Item					10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5			
Israel				++		1	1	0.120
Italy					+	0	1	0.066
Lithuania						0	0	0.061
Malta					–	0	1	0.078
Netherlands		−−	++	++	++	4	4	0.233
New Zealand						0	0	0.027
Northern Ireland	+	++		−−	–	2	4	0.197
Norway			+			0	1	0.087
Poland		+		−−	++	2	3	0.158
Portugal				+		0	1	0.074
Qatar			+			0	1	0.089
Romania				++		1	1	0.082
Russian Federation						0	0	0.073
Saudi Arabia		–	++			1	2	0.133
Singapore			−−		++	2	2	0.104
Slovak Republic						0	0	0.070
Slovenia						0	0	0.059
Spain						0	0	0.056
Sweden	+			−−		1	2	0.090
Trinidad and Tobago						0	0	0.049
United Arab Emirates						0	0	0.060
United States						0	0	0.084

Note + indicates that residual belongs to the 20 % most positive residuals, ++ indicates that residual even belongs to the 10 % most positive residuals. − indicates that residual belongs to the 20 % most negative residuals, −− indicates that residual even belongs to the 10 % most negative residuals. The 10 % cultural differential item functioning (CDIF) and 20 % CDIF columns give the number of outliers in the two respective regions. Absolute residual refers to the means over items of the absolute values of the residuals. Item descriptions are provided in Table 3.2

We then addressed the distribution of country-by-item interaction across countries and items, to determine whether the sizes and directions of the residuals were randomly distributed across all countries and items, or whether they exhibited notable patterns of interaction (Tables 4.21, 4.22, 4.23, 4.24 and 4.25). Residuals were defined by Eq. (4.3), estimated under the GPCM, and calculated for every country, with that country as a focus and all other countries as a reference. To simplify, here we shall not consider the specific values of the residuals, but instead concentrate on the outlying values. For example, if we examine results obtained for the Republic of Azerbaijan and Australia for component 1 (early literacy activities,

Table 4.25 Residual analysis for country-by-item interactions for component 5: school practices on parental involvement, school perspective

Country	Item															10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Azerbaijan					−−											1	1	0.164
Australia								+								0	1	0.152
Austria					++		+			–		+	–			1	5	0.294
Belgium (F)	+		+		–								−−			1	4	0.289
Bulgaria	+														–	0	2	0.227
Canada					+											0	1	0.170
Chinese Taipei							–						++	++		2	3	0.186
Colombia	+											−−		−−		2	3	0.268
Croatia	++				−−			–					++			3	4	0.352
Czech Republic	+					–		+					−−		−−	2	5	0.286
Denmark						–		+				++			++	2	4	0.235
England	+						–						–			0	3	0.209
Finland		−−						++				++	++	++	+	5	6	0.294
France					++	++			+			–	−−			3	5	0.271
Georgia	+											−−		–		1	3	0.195
Germany					++	++										2	2	0.173
Hong Kong					+								++			1	2	0.250
Hungary	++				+	–						−−		−−		3	5	0.237
Indonesia					−−	+	+						−−	++	+	3	6	0.302
Iran					−−	+										1	2	0.203
Ireland							–	++				++		++	+	3	5	0.302
Israel																0	0	0.121

(continued)

Table 4.25 (continued)

Country	Item															10 % CDIF	20 % CDIF	Absolute residual
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Italy	++							–								1	2	0.200
Lithuania												−−				1	1	0.162
Malta													+			0	1	0.124
Netherlands				−−	++							+			+	2	4	0.213
New Zealand					++			++						+	+	2	4	0.247
Northern Ireland												+	–			0	2	0.156
Norway	++				+	–				−−				–		2	5	0.289
Poland					+			++					++			2	3	0.213
Portugal	++		+							–						1	3	0.226
Qatar					–	−−										1	2	0.153
Romania						+								−−	−−	2	3	0.223
Russian Fed.	++					+			–	−−		−−	++	–		4	7	0.367
Saudi Arabia	++				−−			−−		+		–			−−	4	6	0.315
Singapore					+								++			1	2	0.223
Slovak Republic														−−	–	1	2	0.200
Slovenia													++			1	1	0.176
Spain	+							–								0	2	0.176
Sweden									+			++	–			1	3	0.226
Trinidad and Tobago										+		−−	−−	−−		3	4	0.232
United Arab Emirates																0	0	0.128
United States																0	0	0.115

Note + indicates that residual belongs to the 20 % most positive residuals, ++ indicates that residual even belongs to the 10 % most positive residuals. − indicates that residual belongs to the 20 % most negative residuals, −− indicates that residual even belongs to the 10 % most negative residuals. The 10 % cultural differential item functioning (CDIF) and 20 % CDIF columns give the number of outliers in the two respective regions. Absolute residual refers to the means over items of the absolute values of the residuals. Item descriptions are provided in Table 3.2

Table 4.21), it is clear that, aggregated over the items, the mean absolute residual for the Republic of Azerbaijan is much larger than the mean absolute residual for Australia. The responses were coded 0, 1 and 2, so the residuals, which are the differences between a mean observed and expected response are also on a scale from 0 to 2. Closer inspection at the item level for Republic of Azerbaijan reveals that items 3 and 5 have residuals among the 10 % most positive among the countries, while the items 6 and 8 have residuals among the 10 % most negative among the countries. Australia, however, has only one negative residual, and this is among the 20 % most negative residuals among the countries. Checking the absolute residuals further reveals Poland fits the model best with the lowest CDIF, while Indonesia has the most significant CDIF.

In a similar way, component 2 (helping with homework) functions very differently in the Netherlands than in other countries (Table 4.22), probably because giving students homework is not a daily practice in Dutch primary schools. This different item functioning is indicated by both the high mean for the absolute values of the residuals and the large number of outliers among the residuals. Canada fits the model best, having the lowest CDIF for this component. For component 3 (school practices on parental involvement, parents perspective) the highest mean absolute residual was found for Germany. However, the scale for measuring school practices on parental involvement from the school perspective (component 5) showed relatively little evidence of CDIF.

We undertook a marginal count of the outliers for the items aggregated over the countries (Table 4.26). No one item count was prominent, although the first item in component 3 ("my child's school includes me in my child's education") seemed more susceptible to CDIF than other items, since this item had the greatest number of residual outliers among countries: 13 in the 10 % outliers region and 15 in the 20 % outliers region. Items 5 ("volunteering") and 13 ("organize workshops or seminars for parents on learning or pedagogical issues") within component 5 also scored more highly than other items in the component. However, this does not of course mean that these items have CDIF; if 10 and 20 % extreme values are considered, then 10 and 20 % of the residuals must be included, thus such information only serves as a tool to further scrutinize the items.

We also calculated country-specific factor loadings for the bi-factor model, where we first transformed country-specific factor loadings to standard normals, and then identified the 2.5 and 5 % most extreme outlying values (Tables 4.27, 4.28, 4.29, 4.30 and 4.31). This distribution of country-specific factor loadings gives an indication of the extent to which items load on a country-specific factor in addition to the general factor of the item, and can, as in our earlier residual analysis, be used to determine whether the sizes and directions of the factor loadings are randomly distributed across all countries and items, or whether they exhibit notable patterns of interaction.

For component 1, the greatest number of outliers of the country-specific factor loadings and the highest mean absolute factor loading were found for Colombia (Table 4.27), suggesting a high level of CDIF. Interestingly, in the residual analysis for this component, a total of 15 countries showed a higher mean absolute residual

Table 4.26 Distribution of cultural differential item functioning (CDIF) across items on parental involvement

% CDIF	Component	Items														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
10	1	6	2	5	4	6	3	4	5	2	–	–	–	–	–	–
	2	7	2	5	5	1	9	2	2	–	–	–	–	–	–	–
	3	13	0	0	–	–	–	–	–	–	–	–	–	–	–	–
	4	1	2	3	11	5	–	–	–	–	–	–	–	–	–	–
	5	7	1	0	1	10	3	0	5	0	2	0	10	13	9	4
20	1	17	3	7	12	9	5	6	9	5	–	–	–	–	–	–
	2	15	5	8	9	6	12	5	5	–	–	–	–	–	–	–
	3	22	2	1	–	–	–	–	–	–	–	–	–	–	–	–
	4	7	6	8	14	8	–	–	–	–	–	–	–	–	–	–
	5	14	1	2	1	18	11	5	11	3	6	0	15	18	13	11

Note Count of the number of times a residual is in the extreme 10 % and extreme 20 % region of the distribution of residuals. The components, items and corresponding category labels are described in Table 3.2

Table 4.27 Outliers of country-specific factor loadings in the bi-factor model for component 1: early literacy activities

Country	Item									2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3	4	5	6	7	8	9			
Azerbaijan, Republic of					+	−−		−−		3	3	0.062
Australia										0	0	0.034
Austria				−−				−−	−−	3	3	0.071
Belgium (French)	+							++	+	2	3	0.042
Bulgaria										0	0	0.029
Canada								+		1	1	0.050
Chinese Taipei										0	0	0.036
Colombia	−−		−−	–	−−		−−	−−		5	6	0.095
Croatia	−−	−−	–		−−	−−				4	5	0.079
Czech Republic	−−		–	+	−−	−−				4	5	0.083
Denmark	−−				−−	−−				3	3	0.080
Finland	−−	−−								2	2	0.066
France		–			−−					1	2	0.048
Georgia				−−			−−	−−	−−	4	4	0.077
Germany										0	0	0.029
Hong Kong, SAR				−−		−−		−−	−−	4	4	0.082
Hungary		++				–	–	−−		1	4	0.066
Indonesia				++						0	1	0.033
Iran, Islamic Republic of								−−		1	1	0.053
Ireland	−−	−−			–	–				2	4	0.062
Israel	−−	–			–	−−				2	4	0.064
Italy	−−	−−	−−				–			3	4	0.078
Lithuania					–	−−				1	2	0.044

(continued)

Table 4.27 (continued)

Country	Item									2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3	4	5	6	7	8	9			
Malta	−−	–	−−		−−	–				3	5	0.076
Netherlands	++									0	1	0.028
New Zealand	−−		−−		−−	−−				4	4	0.081
Northern Ireland	−−	–		−−	−−				−−	4	5	0.088
Norway	−−		–	–	−−		−−			3	5	0.087
Poland			++	–			−−	−−		2	4	0.056
Portugal										0	0	0.034
Qatar				++						0	1	0.047
Romania				−−			–	−−		2	3	0.060
Russian Federation					++					0	1	0.032
Saudi Arabia	−−	−−				–				2	3	0.065
Singapore				−−			–	−−		2	3	0.057
Slovak Republic	−−				−−	−−				3	3	0.062
Slovenia	−−				−−	−−				3	3	0.065
Spain		–			−−	−−				2	3	0.049
Sweden				−−			–	–		1	3	0.052
Trinidad and Tobago	–	++					–	−−		1	4	0.058
United Arab Emirates	−−	−−			–					2	3	0.065

Note + indicates factor loading belongs to the 5 % most positive loading, ++ indicates factor loading belongs to the 2.5 % most positive loading. − indicates factor loading belongs to the 5 % most negative loading, −− indicates factor loading belongs to the 2.5 % most negative loading. The 2.5 % cultural differential item functioning (CDIF) and 5 % CDIF columns give the number of outliers in the two respective regions. Mean absolute loading refers to the means over items of the absolute values of country-specific factor loadings. Item descriptions are provided in Table 3.2

Table 4.28 Outliers of country-specific factor loadings in the bi-factor model for component 2: help with homework

Country\item	1	2	3	4	5	6	7	8	2.5 % Outlier	5 % Outlier	Mean absolute loading
Azerbaijan, Republic of	+					–			1	2	0.056
Australia						−−	−−	−−	3	3	0.060
Austria								–	0	1	0.037
Belgium (French)	++								0	1	0.043
Bulgaria						−−	−−	−−	3	3	0.051
Canada	+								1	1	0.044
Chinese Taipei						–	−−		1	2	0.037
Colombia	−−			−−		+		–	3	4	0.067
Croatia	−−			–	−−		–	–	2	5	0.068
Czech Republic							–	–	0	2	0.040
Denmark									0	0	0.029
Finland									0	0	0.047
France	+	+				–	–	–	2	5	0.069
Georgia								–	0	1	0.032
Germany	+					–	–		1	3	0.055
Hong Kong, SAR				–				–	0	2	0.044
Hungary	+					−−	−−	–	3	4	0.076
Indonesia						−−	−−	–	2	3	0.057
Iran, Islamic Republic of									0	0	0.048
Ireland									0	0	0.029
Israel	−−			−−				−−	3	3	0.056
Italy									0	0	0.041
Lithuania		++				−−	−−	−−	3	4	0.075

(continued)

Table 4.28 (continued)

Country\item	1	2	3	4	5	6	7	8	2.5 % Outlier	5 % Outlier	Mean absolute loading
Malta	+		+	−−	++		++		3	5	0.072
Netherlands						−−	−−	−−	3	3	0.058
New Zealand						−−	−−	−−	3	3	0.059
Northern Ireland	−−							−−	2	2	0.053
Norway	+					–	−−	−−	3	4	0.066
Poland						−−	−−	–	2	3	0.050
Portugal	++								0	1	0.037
Qatar						−−	−−		2	2	0.041
Romania						−−	−−		2	2	0.051
Russian Federation	+						−−		2	2	0.043
Saudi Arabia				–		+			1	2	0.048
Singapore									0	0	0.036
Slovak Republic								–	0	1	0.043
Slovenia						–	–	–	0	3	0.051
Spain	+					−−			2	2	0.038
Sweden						−−	−−	–	2	3	0.055
Trinidad and Tobago							−−		1	1	0.049
United Arab Emirates								–	0	1	0.036

Note + indicates factor loading belongs to the 5 % most positive loading, ++ indicates factor loading belongs to the 2.5 % most positive loading. − indicates factor loading belongs to the 5 % most negative loading, −− indicates factor loading belongs to the 2.5 % most negative loading. The 2.5 % cultural differential item functioning (CDIF) and 5 % CDIF columns give the number of outliers in the two respective regions. Mean absolute loading refers to the means over items of the absolute values of country-specific factor loadings. Item descriptions are provided in Table 3.2

Table 4.29 Outliers of country-specific factor loadings in the bi-factor model for component 3: school practices on parental involvement, parent perspective

Country	Item			2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3			
Azerbaijan, Republic of	+	++		1	2	1.097
Australia				0	0	0.293
Austria				0	0	0.203
Belgium (French)				0	0	0.223
Bulgaria				0	0	0.262
Canada				0	0	0.423
Chinese Taipei			+	1	1	0.640
Colombia				0	0	0.159
Croatia				0	0	0.194
Czech Republic				0	0	0.393
Denmark				0	0	0.284
Finland				0	0	0.293
France				0	0	0.293
Georgia				0	0	0.240
Germany				0	0	0.409
Hong Kong, SAR				0	0	0.168
Hungary			+	1	1	1.521
Indonesia			++	0	1	0.500
Iran, Islamic Republic of				0	0	0.362
Ireland				0	0	0.279
Israel				0	0	0.216
Italy				0	0	0.131
Lithuania				0	0	0.174
Malta				0	0	0.418
Netherlands				0	0	0.331
New Zealand				0	0	0.260
Northern Ireland				0	0	0.321
Norway				0	0	0.228
Poland				0	0	0.213
Portugal				0	0	0.205
Qatar				0	0	0.297
Romania				0	0	0.430
Russian Federation				0	0	0.153
Saudi Arabia				0	0	0.184
Singapore				0	0	0.150
Slovak Republic				0	0	0.180
Slovenia				0	0	0.228

(continued)

Table 4.29 (continued)

Country	Item			2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3			
Spain				0	0	0.347
Sweden				0	0	0.315
Trinidad and Tobago				0	0	0.517
United Arab Emirates				0	0	0.175

Note + indicates factor loading belongs to the 5 % most positive loading, ++ indicates factor loading belongs to the 2.5 % most positive loading. − indicates factor loading belongs to the 5 % most negative loading, −− indicates factor loading belongs to the 2.5 % most negative loading. The 2.5 % cultural differential item functioning (CDIF) and 5 % CDIF columns give the number of outliers in the two respective regions. Mean absolute loading refers to the means over items of the absolute values of country-specific factor loadings. Item descriptions are provided in Table 3.2

Table 4.30 Outliers of country-specific factor loadings in the bi-factor model for component 4: student perception of parental involvement

Country	Item					2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3	4	5			
Azerbaijan, Republic of						0	0	0.024
Australia						0	0	0.012
Austria						0	0	0.021
Belgium (French)				+		1	1	0.026
Bulgaria						0	0	0.016
Canada	+		++			1	2	0.048
Chinese Taipei						0	0	0.012
Colombia	−−	–				1	2	0.044
Croatia	–	−−	−−	+		3	4	0.084
Czech Republic						0	0	0.018
Denmark						0	0	0.010
England						0	0	0.027
Finland		–				0	1	0.038
France				–	–	0	2	0.035
Georgia	++	++	+	++		1	4	0.057
Germany	−−					1	1	0.034
Hong Kong, SAR						0	0	0.011
Hungary						0	0	0.026
Indonesia				–		0	1	0.029

(continued)

Table 4.30 (continued)

Country	Item					2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3	4	5			
Iran, Islamic Republic of				+		1	1	0.034
Ireland			–			0	1	0.022
Israel						0	0	0.024
Italy						0	0	0.023
Lithuania		−−	++	+		2	3	0.052
Malta						0	0	0.008
Netherlands			–	−−		1	2	0.048
New Zealand						0	0	0.031
Northern Ireland		–				0	1	0.038
Norway						0	0	0.021
Poland				++		0	1	0.025
Portugal				–		0	1	0.030
Qatar			++			0	1	0.034
Romania	++					0	1	0.037
Russian Federation						0	0	0.027
Saudi Arabia			–			0	1	0.028
Singapore						0	0	0.031
Slovak Republic						0	0	0.016
Slovenia						0	0	0.018
Spain						0	0	0.020
Sweden						0	0	0.022
Trinidad and Tobago						0	0	0.020
United Arab Emirates			−−	−−		2	2	0.047
United States						0	0	0.019

Note + indicates factor loading belongs to the 5 % most positive loading, ++ indicates factor loading belongs to the 2.5 % most positive loading. − indicates factor loading belongs to the 5 % most negative loading, −− indicates factor loading belongs to the 2.5 % most negative loading. The 2.5 % cultural differential item functioning (CDIF) and 5 % CDIF columns give the counts of the outliers in the two respective regions. Mean absolute loading refers to the means over items of the absolute values of country-specific factor loadings. Item descriptions are provided in Table 3.2

Table 4.31 Outliers of country-specific factor loadings in the bi-factor model for component 5: school practices on parental involvement, school perspective

Country	Item															2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Azerbaijan																0	0	0.043
Australia								+								1	1	0.078
Austria																0	0	0.048
Belgium (F)	+	+														2	2	0.075
Bulgaria																0	0	0.053
Canada																0	0	0.057
Chinese Taipei	+	+														2	2	0.121
Colombia					−−											1	1	0.066
Croatia																0	0	0.043
Czech Republic																0	0	0.047
Denmark																0	0	0.041
England																0	0	0.041
Finland		+														1	1	0.066
France																0	0	0.028
Georgia													++			0	1	0.047
Germany	+															1	1	0.060
Hong Kong																0	0	0.041
Hungary																0	0	0.070
Indonesia																0	0	0.059
Iran																0	0	0.035
Ireland																0	0	0.050
Israel	+	+														2	2	0.072
Italy																0	0	0.030

(continued)

Table 4.31 (continued)

Country	Item															2.5 % Outlier	5 % Outlier	Mean absolute loading
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Lithuania																0	0	0.035
Malta					+			+								2	2	0.088
Netherlands																0	0	0.041
New Zealand	+	+														2	2	0.116
Northern Ireland	+	+	+													3	3	0.105
Norway										++						0	1	0.083
Poland	+	+	+													3	3	0.113
Portugal	+	+	+	+												4	4	0.179
Qatar																0	0	0.041
Romania																0	0	0.042
Russian Fed.	+	+														2	2	0.067
Saudi Arabia																0	0	0.040
Singapore	++	++														0	2	0.055
Slovak Republic																0	0	0.043
Slovenia										++						0	1	0.063
Spain																0	0	0.044
Sweden																0	0	0.071
Trinidad and Tobago								+								1	1	0.069
United Arab Emirates																0	0	0.054
United States																0	0	0.053

Note + indicates factor loading belongs to the 5 % most positive loading, ++ indicates factor loading belongs to the 2.5 % most positive loading. − indicates factor loading belongs to the 5 % most negative loading, −− indicates factor loading belongs to the 2.5 % most negative loading. The 2.5 % cultural differential item functioning (CDIF) and 5 % CDIF columns give the number of outliers in the two respective regions. Mean absolute loading refers to the means over items of the absolute values of country-specific factor loadings. Item descriptions are provided in Table 3.2

Table 4.32 Distribution of outliers of country-specific factor loadings in the bi-factor model across items on parental involvement

Region	Component	Items														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2.5 %	1	7	6	4	8	13	11	4	12	5	–	–	–	–	–	–
	2	13	1	1	3	1	14	15	8	–	–	–	–	–	–	–
	3	1	0	2	–	–	–	–	–	–	–	–	–	–	–	–
	4	3	2	3	6	0	–	–	–	–	–	–	–	–	–	–
	5	9	9	3	1	2	0	0	3	0	0	0	0	0	0	0
5 %	1	19	8	13	13	18	15	10	14	5	–	–	–	–	–	–
	2	15	2	1	6	2	20	21	22	–	–	–	–	–	–	–
	3	1	1	3	–	–	–	–	–	–	–	–	–	–	–	–
	4	6	6	9	11	1	–	–	–	–	–	–	–	–	–	–
	5	10	10	3	1	2	0	0	3	0	2	0	0	1	0	0

Note Count of the number of times an outlier of the country-specific factor loading is in the extreme 2.5 % and extreme 5 % region of the distribution of factor loadings. The components, items and corresponding category labels are described in Table 3.2

Table 4.33 Relation between residuals under the generalized partial credit model (GPCM) and country-specific factor loadings in the bi-factor GPCM

Component	Correlation	Kappa classification CDIF		
		Size middle group		
		33 %[a]	40 %[b]	80 %[c]
1	0.228	0.15	0.20	0.24
2	0.603	0.21	0.29	0.27
3	−0.044	0.07	0.17	0.10
4	0.651	0.46	0.41	0.41
5	0.519	0.34	0.31	0.25

Note Correlation between the GPCM residuals and the country-specific factor loadings, over countries and items. The content of the components is described in Table 3.2. Size middle group indicates the classification of the ordered residuals and country-specific factor loadings in three categories according to their size: a category with negative values, a category with positive values and a middle category

Norms for Kappa: poor agreement = 0.00–0.19, fair agreement = 0.20–0.39, and moderate agreement = 0.40–0.59

[a]Three equally-sized categories

[b]The middle category contained 40 % of the values, the two extreme categories each contained 20 %

[c]The middle category contained 80 % of the values, the two extreme categories each contained 10 %

(Table 4.21). Regarding help with homework (component 2), Malta was identified as having the highest number of outliers in country-specific factor loadings (Table 4.28), while The Netherlands, which we earlier identified as exhibiting CDIF for component 2 (Table 4.22), also had a high number of outliers. For component 3, counting the number of outliers provided little information, as only three outliers were counted in the 2.5 % region (Table 4.29). Hungary did show a high mean absolute country-specific factor loading on this component, though the questionable reliability of the scale must be kept in mind. Student perception of parental involvement (component 4) was measured with the least CDIF in Denmark, whereas the school practices on parental involvement from the school perspective showed the least CDIF for Italy (Tables 4.30 and 4.31).

Aggregating the items over the countries provides a tool for further investigation of items (Table 4.32), with the same caveats as before; if the 2.5 and 5 % most extreme values are considered, then similarly 2.5 and 5 % of the residuals must fall in this region, but this does not imply that 2.5 and 5 % of the items have CDIF. No item count is prominent. Item 5 ("talk about things you had done") in component 1 did seem more susceptible to CDIF than other items, since this item revealed the greatest number of outliers in country-specific factor loadings over countries.

We then addressed whether the residual analyses using the GPCM and the bi-factor GPCM analyses led to the same conclusions (see Table 4.33). A priori, this would be unexpected. The residual analyses target so-called uniform CDIF, namely a shift in the item location (item intercept) parameters over countries. The

Table 4.34 Correlation and rank correlation between country means estimated with no, 10 and 20 % cultural differential item functioning (CDIF) parameters, and random item parameters

Component	Parameter	Correlation			Rank correlation		
		No CDIF	10 % CDIF	20 % CDIF	No CDIF	10 % CDIF	20 % CDIF
1	10 % CDIF	0.99	–	–	0.98	–	–
	20 % CDIF	0.99	0.99	–	0.98	0.98	–
	Random	0.98	0.97	0.97	0.97	0.96	0.97
2	10 % CDIF	0.99	–	–	0.98	–	–
	20 % CDIF	0.98	0.99	–	0.98	0.99	–
	Random	0.66	0.64	0.58	0.95	0.93	0.95
3	10 % CDIF	0.83	–	–	0.94	–	–
	20 % CDIF	0.80	0.82	–	0.93	1.00	–
	Random	0.53	0.38	0.33	0.62	0.64	0.63
4	10 % CDIF	0.98	–	–	0.97	–	–
	20 % CDIF	0.97	0.98	–	0.95	0.95	–
	Random	0.50	0.44	0.37	0.94	0.92	0.89
5	10 % CDIF	0.97	–	–	0.97	–	–
	20 % CDIF	0.97	1.00	–	0.97	1.00	–
	Random	0.97	0.98	0.98	0.97	0.99	0.99

Note The components are described in Table 3.2

bi-factor analyses target non-uniform CDIF, namely differences in the slopes and the dimensionality across items. The correlations for components 2, 4 and 5 were moderate, while for component 1, the correlation was much lower, and for component 3, the correlation completely vanished. The result for component 3 is probably because both the residuals and the country-specific factor loadings are poorly estimated for a test containing only three items.

Though the correlation between the residuals and the country-specific factor loadings is a reasonable estimate between the two measures, it does not properly indicate to what extent the two measures have the same outliers. To investigate this, we ordered and classified the residuals and country-specific factor loadings in three categories according to their size (a category with negative values, a category with positive values and a middle category). Further, we varied the definition of what constituted an outlying value by varying the size of the middle group (assigning it variously as 33, 40, or 80 % of values). The calculation of Kappa establishes the agreement in categorization between the residual analyses using the GPCM and the bi-factor GPCM. This revealed that agreement was poor throughout for component 3, while, for component 1, the agreement was poor in the 33 % category; for other categories in component 1 the agreement was only fair to moderate. In general, the results indicate that it is not a good policy to rely on one approach for the investigation of CDIF.

We investigated the influence of CDIF by calculating the correlation and rank correlation between country means estimated with no, 10, and 20 % CDIF

parameters, and with random item parameters (Table 4.34). Estimates of the means using the unidimensional GPCM without country-specific item parameters and using the bi-factor GPCM could not be distinguished, so we exclude them from further discussion. In general, correlations were high, indicating that, in the estimation of the country means and the rank order of the country means, CDIF had little impact. Component 3 remained the exception; both correlations and rank correlations were low. Further, for components 2 and 4, the correlations between the means estimated using the GPCM with random item parameters and the other three models were also low; however this was not the case for the rank correlations. This is because the relationship between means is not linear. We discuss the possible influence of CDIF further in the next chapter.

References

Adams, R., & Wu, M. (2006). The mixed-coefficients multinomial logit model: A generalized form of the Rasch model. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 57–75). New York: Springer.

Aitchison, J., & Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics, 29*, 813–828.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM-algorithm. *Psychometrika, 46*, 443–459.

Bock, R. D., Gibbons, R. D., & Muraki, E. (1988). Full-information factor analysis. *Applied Psychological Measurement, 12*, 261–280.

De Jong, M. G., Steenkamp, J. B. E. M., & Fox, J. P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research, 34*(2), 260–278. doi:10.1086/518532.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*(3), 423–436. doi:10.1007/BF02295430.

Glas, C. A. W. (1999). Modification indices for the 2-PL and the nominal response model. *Psychometrika, 64*(3), 273–294. doi:10.1007/bf02294296.

Glas, C. A. W. (2010). Multidimensional item response theory (MIRT), manual and computer program. Retrieved from http://www.utwente.nl/gw/omd/Medewerkers/temp_test/mirt_package.zip, http://www.utwente.nl/gw/omd/Medewerkers/temp_test/mirt-manual.pdf.

Glas, C. A. W., & Dagohoy, A. V. T. (2007). A person fit test for IRT models for polytomous items. *Psychometrika, 72*(2), 159–180. doi:10.1007/s11336-003-1081-5.

Glas, C. A. W., & Falcón, J. C. S. (2003). A comparison of item-fit statistics for the three-parameter logistic model. *Applied Psychological Measurement, 27*(2), 87–106. doi:10.1177/0146621602250530.

Glas, C. A. W., & Jehangir, K. (2014). Modeling country-specific differential item functioning. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 97–115). New York: Springer.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 76*(4), 537–549. doi:10.1007/s11336-011-9218-4.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159–176.

Rao, C. R. (1947). Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society, 44*, 50–57.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph, 34*(4, Pt. 2), 1–100.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology, 43*, 39–55.