NEPS SURVEY PAPERS

Nicolas Hübner, Sven Rieger, and Wolfgang Wagner

NEPS TECHNICAL REPORT FOR BIOLOGICAL COMPETENCE: SCALING RESULTS FOR THE ADDITIONAL STUDY BADEN-WUERTTEMBERG

LIfBi

**LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES**

NEPS
National Educational Panel Study

Survey Papers of the German National Educational Panel Study (NEPS)
at the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg

The NEPS Survey Paper Series provides articles with a focus on methodological aspects and data handling issues related to the German National Educational Panel Study (NEPS).

The NEPS Survey Papers are edited by a review board consisting of the scientific management of LIfBi and NEPS.

They are of particular relevance for the analysis of NEPS data as they describe data editing and data collection procedures as well as instruments or tests used in the NEPS survey. Papers that appear in this series fall into the category of 'grey literature' and may also appear elsewhere.

The NEPS Survey Papers are available at https://www.neps-data.de (see section "Publications").

Editor-in-Chief: Corinna Kleinert, LIfBi/University of Bamberg/IAB Nuremberg

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg – Germany – contact@lifbi.de

NEPS Technical Report for Biological Competence:

Scaling Results for the

Additional Study Baden-Wuerttemberg

Nicolas Hübner, Sven Rieger, & Wolfgang Wagner

Hector Research Institute of Education Sciences and Psychology,
University of Tübingen

E-mail address of lead author:

nicolas.huebner@uni-tuebingen.de

NEPS Technical Report for Biological Competence: Scaling Results for the Additional Study Baden-Wuerttemberg

Abstract

The National Educational Panel Study (NEPS) is aimed at investigating the development of competences across the entire life span. It also develops tests for assessing different competence domains. In order to evaluate the quality of these competence tests, a wide range of item response theory (IRT) analyses were carried out. This paper describes the data and results of analyses of the biological competence test that was used in the additional study Baden-Wuerttemberg. The test was based on a subset of items from the EVAMAR II test (Eberle, et al. 2008), which was also administered in the additional study Thuringia. In sum, 4,882 students took the biology test in these three waves. The biology test consisted of 60 items, representing different content areas (e.g. cytology, genetics, ecology). A partial credit model was used to scale the data. Item fit statistics and differential item functioning were investigated. The results showed that the items exhibited good item fit and measurement invariance across various groups. However, the reliability was moderate. The paper also provides some information about the data available in the Scientific Use File, ConQuest- and TAM-syntaxes for scaling the data, and appendices that describe the scaling of each wave separately.

Keywords

item response theory, scaling, biological competence, Scientific Use File

Contents

1. Introduction

In the National Educational Panel Study (NEPS), different competences are measured coherently across the life span. Tests have been developed for different competence domains. These include, among other things, reading competence, mathematical competence, scientific literacy, information and communication technologies literacy, metacognition, vocabulary, and domain-general cognitive functioning. Wagner et al. (2011) provide an overview of the stage 5 competences measured in NEPS.

Most of the competence data are scaled with models that are based on item response theory (IRT). Because most of the competence tests were developed specifically for implementation in NEPS, several analyses have been conducted to evaluate the quality of the tests. The IRT models chosen to scale the competence data and the analyses performed to check the quality of the scales are described in Pohl and Carstensen (2012).

This paper presents the results of the biological competence test in three waves of the additional study Baden-Wuerttemberg. Items in this test were developed in the Swiss EVAMAR II project (Eberle et al., 2008) and used across three consecutive years (2011 through 2013) to test secondary-school students' biological competences in their final year of Gymnasium (the type of school that leads to upper secondary education and the Abitur). More detailed information about the aims of this study can be found on the NEPS website.[1] Further information about the test can be found in NEPS (2011; 2012).

The present report draws strongly on previous technical reports such as Hübner, Rieger, & Wagner (2016), Durchhardt (2015), Pohl, Haberkorn, Hardt, and Wiegand (2012) and Pohl and Carstensen (2012). It includes extracts from these previous reports.

2. Testing Biological Competence

The biological competence test was originally developed in the course of the Swiss EVAMAR project (Evaluation der Maturitätsreform) and is based on a model, which links cognitive dimensions with specific content areas. The content-related aspects are based on the analysis of four German standard biology textbooks for upper secondary school (Biologie Oberstufe, 2001; Biologie heute entdecken, 2004; Linder Biologie, 2005; Natura, 2006). Furthermore, the cognitive aspects are oriented at the EPA standards (Einheitliche Prüfungsanforderung in der Abiturprüfung Biologie) for the final examinations in biology in upper secondary school (KMK, 2004). In the following, we will outline specific aspects of the biological competence paper-and-pencil test that are necessary for understanding the scaling results presented in this paper. The items are not arranged in units. Thus, on the test, students must usually read descriptions of a certain situation and have to answer only one task related to.

There are two types of response formats on the biological competence test. These are single multiple choice (MC), and short constructed response (SCR; see Table 2). For MC items, the

---

[1] https://www.neps-data.de/en-us/datacenter/studydocumentation/additionalstudybadenwuerttemberg.aspx

test taker has to choose the correct answer out of several - usually four - response options. SCR items require the test taker to write down an answer into an empty field.

Tables 1 and 2 show how the content areas and response formats are distributed across the items.

Table 1

Content Areas of the Items on the Biology Test

Content area	Frequency
Cytology/ anatomy/ metabolism	27
Information processing/ characteristics/ immunobiology	10
Genetics/ developmental biology	7
Ecology	11
Systematics/ evolution	5
Total number of items	60

Table 2

Response Formats of the Items on the Biology Test

Response format	Frequency
Single multiple choice	56
Short constructed response	4
Total number of items	60

3. Data

A description of the design of the study, the sample, as well as the instruments that were used can be found on the NEPS website.[2] A total of 4,882 participants took the biology competence test: 1,283 in 2011 (Wave 1), 2,387 in 2012 (Wave 2), and 1,212 in 2013 (Wave 3). All subjects gave at least one valid answer so that for every subject, one competence score was estimated.

4. Analyses

This section briefly describes the analyses that were computed; these included inspecting the various missing responses, scaling the data, and examining the psychometric quality of the test.

4.1 Missing Responses

There are different types of missing responses in competence test data. These include missing responses due to a) invalid responses, b) omitted items, c) items that test takers did not reach, and d) items that are missing by design. Missing responses provide information about how well the test worked (e.g., time limits, whether participants understood the instructions, how participants handled different response formats), and they need to be accounted for in the estimation of item and person parameters. We thoroughly inspected the occurrence of missing responses per person. This provided an indication of how well the test takers coped with the test. We then examined the occurrence of missing responses per item in order to obtain some information about how well the items performed. In addition, information was available about whether students did not take the biological competence test (e.g., due to student tardiness) but did take at least one of the other competence tests (mathematics, English, or physics). This missing code is referred to as e) missing by non-participation.

4.2 Scaling Model

In order to estimate the item and person parameters for biological competence, a 1PL partial credit model (PCM; Masters, 1982) was used and estimated in ConQuest 4.2.5 (Wu, Adams, & Wilson, 1997).

Item parameters are estimated difficulties for dichotomous variables and location parameters for polytomous variables in the PCM. Ability estimates for biological competence were estimated as weighted maximum likelihood estimates (WLEs; Warm, 1989). Person parameter estimation in NEPS is described by Pohl and Carstensen (2012a), whereas the data available in the SUF are described in Section 7.

Plotting the item parameters in relation to the ability estimates of the persons was used in order to judge how well the item difficulties were targeted toward the test persons' abilities (see Figure 5). The test targeting provides some information about the precision of the ability estimates at different levels of ability.

---

2 https://www.neps-data.de/de-de/datenzentrum/datenunddokumentation/zusatzstudiebaden-w%C3%BCrttemberg/dokumentation.aspx

4.3   Checking the Quality of the Scale

To ensure that the test featured appropriate psychometric properties, the quality of the test was examined with several analyses.

The item fit was examined using a partial credit model (Masters, 1982) an extension of the Rasch model for polytomous items. We examined the weighted (or "infit") mean square (WMNSQ), the respective t-value, and correlations between the item scores and the total score. In accordance with Pohl and Carstensen (2012), items with a WMNSQ > 1.15 (t-value > |6|) were considered to have a noticeable item misfit, and items with a WMNSQ > 1.20 (t-value > |8|) were considered to have a considerable item misfit, and their performance was further investigated. Correlations between an item score and the total score (equal to the discrimination as computed in ConQuest) greater than 0.3 were considered good, greater than 0.2 acceptable, and below 0.2 problematic. Overall, the judgment of item fit was based on all fit indicators.

Our aim was to consider a biology competence that measured the same construct in all participants. If any items favored a certain subgroup (e.g., items that were easier for males than for females), measurement invariance would be violated, and a comparison of competence scores between the subgroups (e.g., males and females) would be biased and thus unfair.[3] We addressed the issue of measurement invariance by investigating test fairness for the variables gender, immigration background, books at home (as a proxy for socioeconomic status), and wave (i.e., to which of the three waves do subjects belong?); see Pohl and Carstensen (2012) for a description of these variables. Differential item functioning (DIF) was estimated by applying a multifaceted IRT model in ConQuest in which the main effects of the subgroups and the differential effects of the subgroups on item difficulty were modeled. Differences in the estimated item difficulties between the subgroups were evaluated. On the basis of our experiences with the preliminary data (e.g., Pohl & Carstensen, 2012), we judged absolute differences in estimated difficulties that were greater than 1 logit as having very strong DIF, absolute differences between 0.6 and 1 as worthy of further investigation, differences between 0.4 and 0.6 as considerable but not significant, and differences smaller than 0.4 as not having any considerable DIF. In addition to computing DIF analyses at the item level, we investigated test fairness by comparing a model that included differential item functioning with a model that estimated only main effects but no DIF.

The biology competence data were scaled with the partial credit model (1PL), which assumes Rasch homogeneity. Nonetheless, Rasch homogeneity is an assumption that might not hold for empirical data. We therefore checked for deviations from uniform discrimination. We estimated item discrimination by applying the generalized partial credit model (2PL) (Muraki, 1990) with the TAM package in R (Kiefer, Robitzsch, & Wu, 2015; R Core Team, 2016).

---

[3] It should be noted that differential item functioning may also reflect valid differences between subgroups – that is, item impact (Zumbo, 1999).

5. Results

In this section, the key scaling results of the three waves of the additional study Baden-Wuerttemberg will be presented. Some results in which each wave was scaled separately can be found in Appendices C1–C3.

5.1 Missing Responses

In this subsection, we first report the number of missing responses that can be categorized into the different types of missing responses as described in Chapter 4.1 per person and the total number of missing responses per person. Afterwards, we describe the missing responses per item.

5.1.1 Missing responses per person

Figure 1 shows the number of *invalid responses* per person. As can be seen, 83.0 % of the participants did not produce any invalid responses. The maximum number of invalid responses was 21. Overall, 2.64% of the participants produced five or more invalid responses.
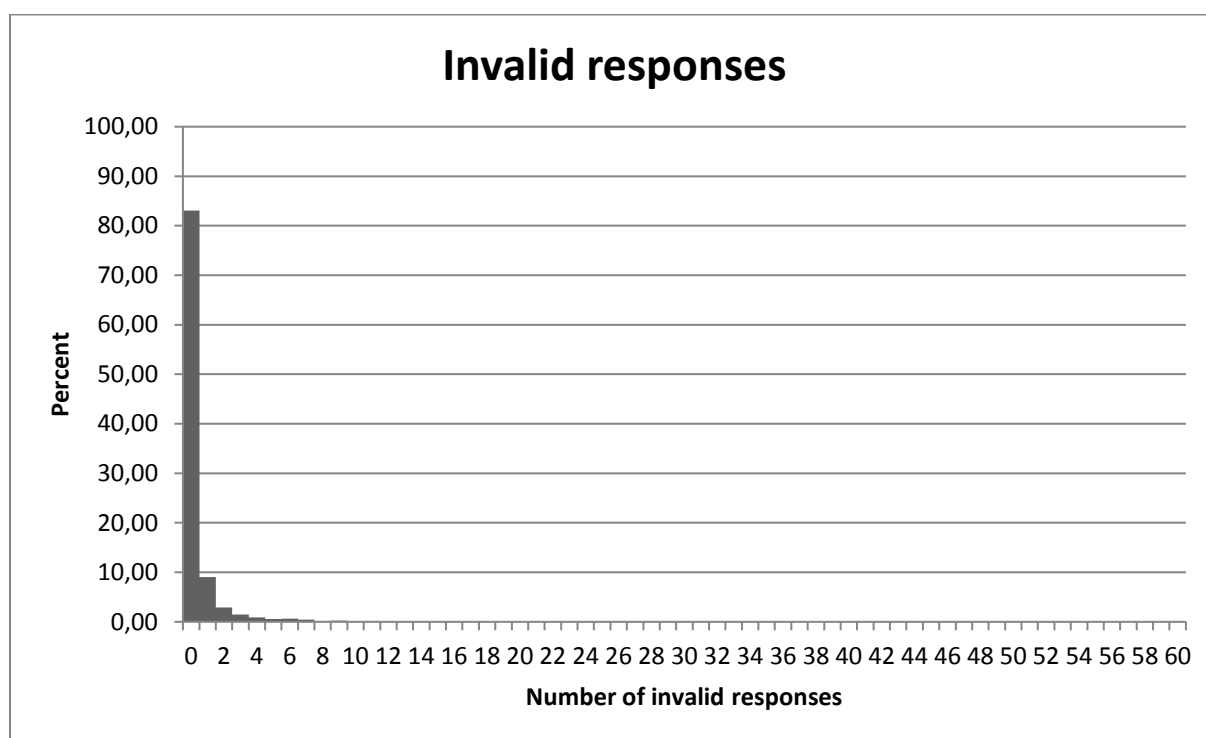
Figure 1. Number of invalid responses per person.

The largest source of missing responses on this test was the *omission of items*. As can be seen in Figure 2, more than half of the participants (52.0%) skipped at least one item. 6.1% of the participants omitted five or more items.
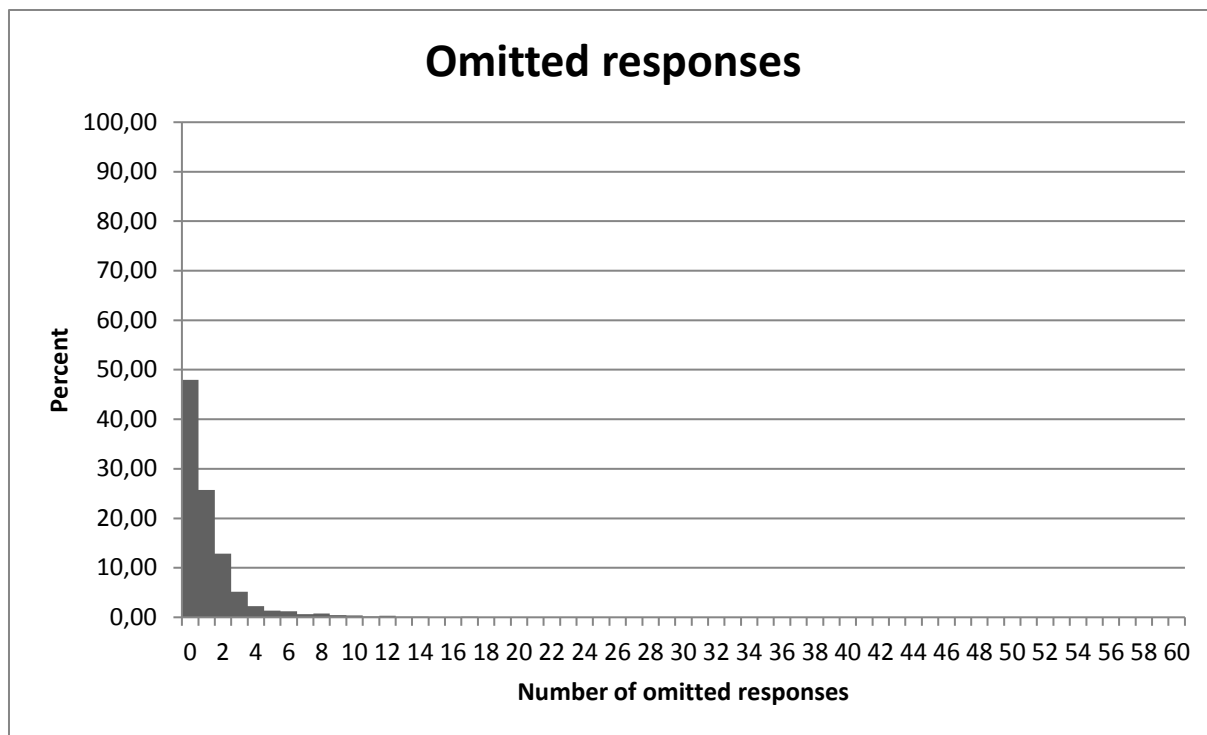
Figure 2. Number of omitted responses per person.

By definition, every item after the last item that was completed is labeled *not reached*. As Figure 3 shows, most participants (94.9%) reached the end of the test. Only 0.2% did not reach the fifth last item.
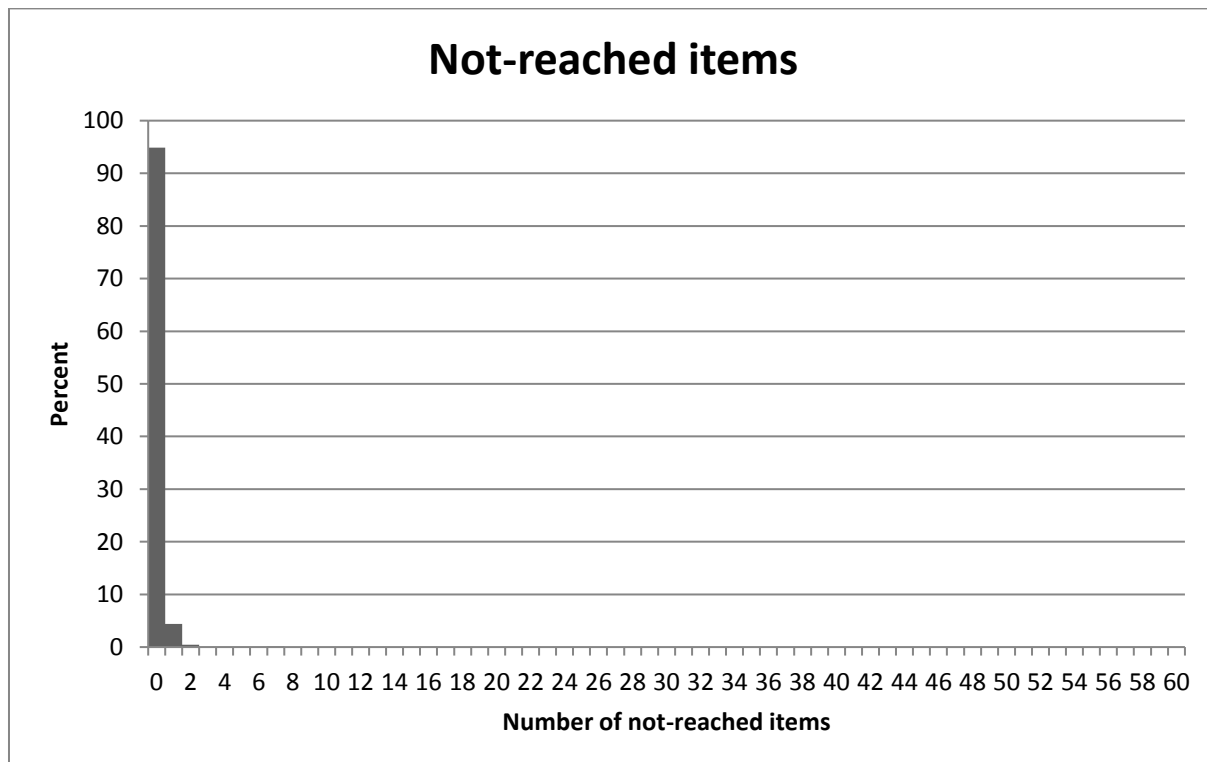
Figure 3. Number of not-reached items per person.

Overall, 99.8% of the participants had no items that were missing by *non-participation*. Only 0.2% (11 of the students) did not take the biological competence test but did take at least one of the other tests.

The total number of missing responses (excluding those missing by non-participation and missing by design) aggregated across the invalid, omitted, and not-reached missing responses per person is illustrated in Figure 4. On average, the participants produced 1.79 (SD = 2.9) missing responses. Moreover, 39.5% of the participants had no missing responses at all (except missing by design). Only 10.3% of the participants had five or more missing responses.
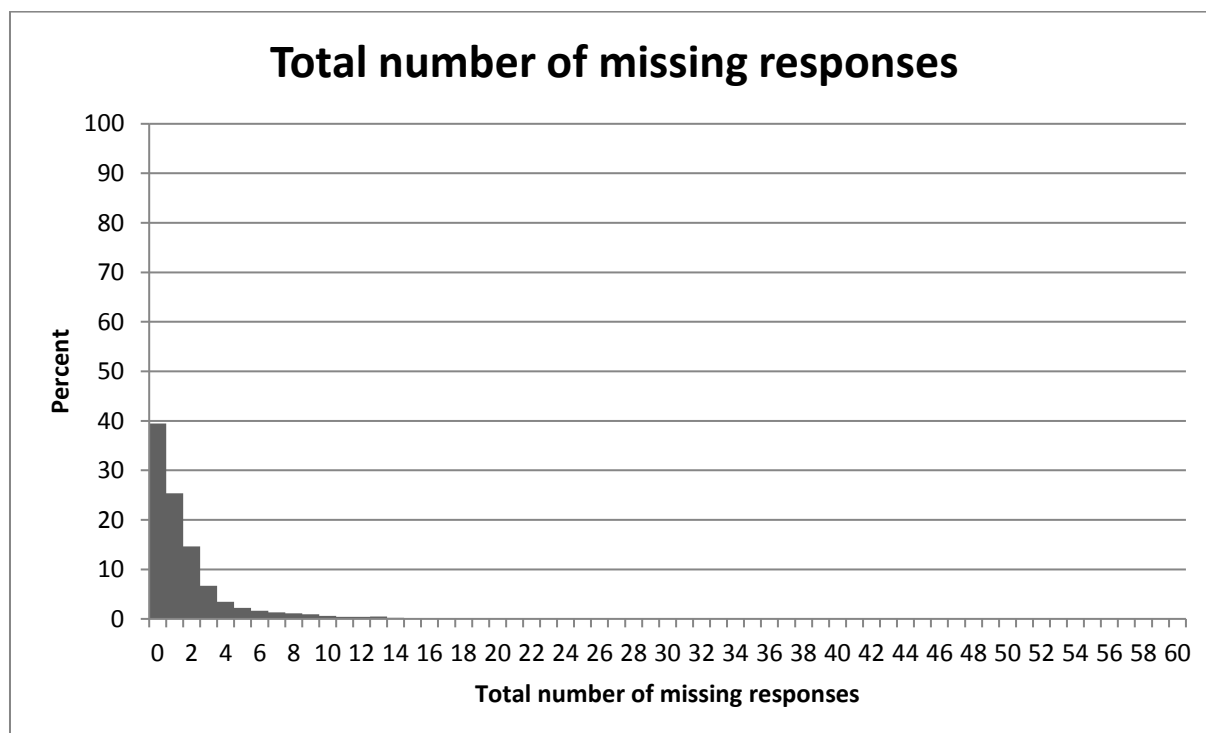
Figure 4. Total number of missing responses.

5.1.2 Missing responses per item

Table 3 provides information about the occurrence of the different kinds of responses that were missing per item. A maximum of 3.1% of the participants failed to reach items (column 5). However, with about 95% of the participants reaching the end of the test, the test can hardly be described as too long. Overall, 5 out of the 60 items had omission rates exceeding 5% (column 6). Items with open response format (short constructed responses) binf21_c (omitted by 22.8% of the participants), bstw21_c (22.1%) and bfkt21_c (14.6%) were the most noticeable. Overall, the percentage of invalid responses per item (column 7) was low (the maximum was 6.3% for item bevo20_c). The percentage of items that were missing by non-participation (column 8) was very low (the maximum was 0.2%). Controlling for students who did not participate in the test, but in other tests, all students had 24 items that were missing by design (column 9).

5.2   Parameter Estimates

5.2.1 Item parameters

The second column in Table 4 shows the percentage of correct responses relative to all valid responses for each item. Please note that, because there is a nonnegligible number of missing responses, this probability cannot be interpreted as an index of item difficulty. The percentage of correct responses varied from 17.0% to 93.2% with an average of 58.22% (SD = 18.86%) correct responses.

For reasons of model identification, in the partial credit model the mean of the ability distribution was constrained to be zero. The estimated item difficulties are given in the third column of Table 4. The item difficulties ranged from -2.80 (item boek19_c) to 5.98 (item bstw21_c) logits with an average difficulty of -0.28 logits (SD = 1.30). Altogether, the item difficulties were adequate. Owing to the large sample size, the corresponding standard errors of the estimated item difficulties (column 4) were small (SE(ß) ≤ 0.079), except for item bstw21_c (SE(ß) = 0.502). The extreme values for bstw21_c resulted from the fact that only very few people were able to correctly respond to the second part of the polytomous item. The step parameters for the polytomous items are displayed in Table 5.

Table 3

Item Parameters of the Biology Test

	Item	Booklet	Position in the test	Number of valid responses	Percentage of not-reached responses	Percentage of omitted responses	Percentage of invalid responses	Percentage of missing by non-participation	Percentage of missing by design
1	bevo16_c	1,2,3,4	1	4806	-	0.6	1.0	0.2	-
2	bfkt06_c	1,2,3,4	2	4749	-	2.1	0.7	0.2	-
3	bfkt16_c	1,2,3,4	3	4583	-	6.1	0.0	0.2	-
4	bfkt18_c	1,2,3,4	4	4794	-	1.7	0.1	0.2	-
5	bgen08_c	1,2,3,4	5	4727	-	1.9	1.3	0.2	-
6	binf10_c	1,2,3 ,4	6	4852	-	0.6	0.1	0.2	-
7	binf21_c	1,2,3,4	7	3721	-	22.8	1.0	0.2	-
8	boek17_c	1,2,3,4	8	4763	-	1.0	1.4	0.2	-
9	boek19_c	1,2,3,4	9	4815	-	0.1	1.2	0.2	-
10	bstw07_c	1,2,3,4	10	4767	-	1.8	0.6	0.2	-
11	bstw12_c	1,2,3,4	11	4830	-	0.8	0.3	0.2	-
12	bstw19_c	1,2,3,4	12	4816	-	1.2	0.2	0.2	-
13	bevo20_c	1,2	25 / 13	2079	0.0	1.2	6.3	0.2	49.7
14	bfkt10_c	1,2	26 / 14	2387	0.0	1.2	0.1	0.2	49.7

	Item	Booklet	Position in the test	Number of valid responses	Percentage of not-reached responses	Percentage of omitted responses	Percentage of invalid responses	Percentage of missing by non-participation	Percentage of missing by design
15	bfkt13_c	1,2	27 / 15	2312	0.0	2.7	0.1	0.2	49.7
16	bfkt14_c	1,2	28 / 16	2294	0.0	0.7	2.5	0.2	49.7
17	bgen14_c	1,2	29 / 17	2368	0.0	1.0	0.7	0.2	49.7
18	binf09_c	1,2	30 / 18	2314	0.0	0.8	1.9	0.2	49.7
19	binf11_c	1,2	31 / 19	2376	0.0	0.6	0.9	0.2	49.7
20	boek04_c	1,2	32 / 20	2198	0.0	5.0	0.1	0.2	49.7
21	boek10_c	1,2	33 / 21	2376	0.0	0.6	0.9	0.2	49.7
22	bstw08_c	1,2	34 / 22	2411	0.0	0.2	0.6	0.2	49.7
23	bstw17_c	1,2	35 / 23	2371	0.3	1.0	0.3	0.2	49.7
24	bstw21_c	1,2	36 / 24	1164	3.1	22.1	1.2	0.2	49.7
25	bevo07_c	1,4	13 / 25	2179	-	1.9	1.7	0.2	51.7
26	bfkt02_c	1,4	14 / 26	2303	-	0.2	0.8	0.2	51.7
27	bfkt04_c	1,4	15 / 27	2266	-	0.6	1.2	0.2	51.7
28	bfkt07_c	1,4	16 / 28	2297	-	0.8	0.4	0.2	51.7
29	bgen04_c	1,4	17 / 29	2294	-	1.1	0.1	0.2	51.7
30	bgen18_c	1,4	18 / 30	2258	-	1.8	0.2	0.2	51.7

	Item	Booklet	Position in the test	Number of valid responses	Percentage of not-reached responses	Percentage of omitted responses	Percentage of invalid responses	Percentage of missing by non-participation	Percentage of missing by design
31	binf08_c	1,4	19 / 31	2254	-	2.0	0.0	0.2	51.7
32	binf17_c	1,4	20 / 32	2272	0.0	1.1	0.6	0.2	51.7
33	boek03_c	1,4	21 / 33	2338	0.0	0.2	0.1	0.2	51.7
34	boek09_c	1,4	22 / 34	2231	0.1	2.3	0.1	0.2	51.7
35	bstw04_c	1,4	23 / 35	2308	0.2	0.6	0.1	0.2	51.7
36	bstw18_c	1,4	24 / 36	2312	0.5	0.2	0.1	0.2	51.7
37	bevo01_c	2,3	25 / 13	2491	-	0.3	0.5	0.2	48.1
38	bfkt09_c	2,3	26 / 14	2497	-	0.6	0.1	0.2	48.1
39	bfkt17_c	2,3	27 / 15	2433	-	1.4	0.5	0.2	48.1
40	bfkt19_c	2,3	28 / 16	2477	-	0.5	0.5	0.2	48.1
41	bgen05_c	2,3	29 / 17	2453	-	1.2	0.3	0.2	48.1
42	bgen11_c	2,3	30 / 18	2439	-	0.8	1.0	0.2	48.1
43	binf07_c	2,3	31 / 19	2475	0.0	0.4	0.6	0.2	48.1
44	binf19_c	2,3	32 / 20	2430	0.0	1.7	0.3	0.2	48.1
45	boek16_c	2,3	33 / 21	2462	0.0	0.9	0.5	0.2	48.1
46	boek20_c	2,3	34 / 22	2443	0.0	1.0	0.7	0.2	48.1

	Item	Booklet	Position in the test	Number of valid responses	Percentage of not-reached responses	Percentage of omitted responses	Percentage of invalid responses	Percentage of missing by non-participation	Percentage of missing by design
47	bstw13_c	2,3	35 / 23	2490	0.2	0.4	0.1	0.2	48.1
48	bstw16_c	2,3	36 / 24	2492	0.3	0.1	0.3	0.2	48.1
49	bevo02_c	3,4	25 / 13	2411	-	0.3	0.1	0.2	50.1
50	bfkt03_c	3,4	26 / 14	2396	-	0.1	0.7	0.2	50.1
51	bfkt21_c	3,4	27 / 15	1697	-	14.6	0.5	0.2	50.1
52	bgen21_c	3,4	28 / 16	2058	-	7.5	0.2	0.2	50.1
53	binf05_c	3,4	29 / 17	2340	-	0.1	1.8	0.2	50.1
54	binf20_c	3,4	30 / 18	2267	0.1	1.0	2.2	0.2	50.1
55	boek02_c	3,4	31 / 19	2350	0.1	0.4	1.1	0.2	50.1
56	boek12_c	3,4	32 / 20	2366	0.1	1.0	0.2	0.2	50.1
57	boek13_c	3,4	33 / 21	2343	0.2	1.6	0.1	0.2	50.1
58	bstw06_c	3,4	34 / 22	2337	0.2	1.0	0.8	0.2	50.1
59	bstw09_c	3,4	35 / 23	2392	0.2	0.1	0.6	0.2	50.1
60	bstw11_c	3,4	36 / 24	2274	1.0	1.0	1.2	0.2	50.1

Table 4

Item Parameters of the Biology Test

	Item	Percentage cor-rect	Difficulty/ loca-tion parameter	*SE* (difficulty/ loca-tion parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination-2 PL
1	bevo16_c	58.80	-0.394	0.032	1.00	-0.5	0.36	0.71
2	bfkt06_c	42.28	0.340	0.032	1.00	0.4	0.34	0.64
3	bfkt16_c	36.53	0.605	0.034	1.06	5.1	0.24	0.34
4	bfkt18_c	66.65	-0.762	0.033	1.09	7.3	0.17	0.18
5	bgen08_c	49.42	0.019	0.032	1.04	4.4	0.28	0.44
6	binf10_c	86.25	-1.986	0.044	1.02	0.6	0.21	0.54
7	binf21_c	n.a.	1.313	0.050	0.94	-3.2	0.40	0.63
8	boek17_c	22.78	1.332	0.037	0.95	-2.6	0.39	0.97
9	boek19_c	93.23	-2.803	0.059	0.99	-0.1	0.19	0.70
10	bstw07_c	48.56	0.063	0.032	0.99	-1.3	0.38	0.77
11	bstw12_c	82.07	-1.651	0.040	0.96	-1.8	0.36	1.08
12	bstw19_c	67.86	-0.821	0.034	1.00	-0.2	0.34	0.72
13	bevo20_c	44.11	0.250	0.048	1.00	0.3	0.34	0.60
14	bfkt10_c	54.04	-0.188	0.044	0.98	-1.6	0.39	0.77
15	bfkt13_c	39.88	0.444	0.046	0.99	-0.8	0.36	0.72

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
16	bfkt14_c	74.76	-1.189	0.051	1.03	1.2	0.26	0.49
17	bgen14_c	71.11	-0.992	0.049	1.03	1.5	0.27	0.51
18	binf09_c	36.39	0.598	0.047	1.04	2.5	0.27	0.43
19	binf11_c	73.57	-1.131	0.050	0.98	-0.7	0.34	0.73
20	boek04_c	52.68	-0.124	0.046	1.03	2.7	0.29	0.47
21	boek10_c	70.75	-0.970	0.048	1.01	0.6	0.30	0.62
22	bstw08_c	76.19	-1.279	0.051	1.03	1.0	0.25	0.48
23	bstw17_c	65.42	-0.706	0.046	0.96	-2.6	0.41	1.00
24	bstw21_c	n.a.	5.980	0.502	0.99	-0.1	0.20	0.85
25	bevo07_c	51.63	-0.080	0.046	1.02	1.3	0.32	0.55
26	bfkt02_c	58.01	-0.372	0.046	1.01	0.4	0.34	0.65
27	bfkt04_c	51.90	-0.104	0.046	1.03	2.6	0.30	0.55
28	bfkt07_c	37.92	0.521	0.046	0.97	-2.3	0.39	0.87
29	bgen04_c	38.97	0.474	0.046	1.04	2.7	0.27	0.46
30	bgen18_c	16.96	1.704	0.059	1.07	1.8	0.13	0.19
31	binf08_c	58.07	-0.376	0.046	0.99	-0.5	0.36	0.71
32	binf17_c	63.25	-0.613	0.047	1.00	-0.1	0.35	0.69

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
33	boek03_c	81.61	-1.636	0.056	0.98	-0.5	0.31	0.81
34	boek09_c	46.08	0.151	0.046	1.00	-0.3	0.36	0.68
35	bstw04_c	84.84	-1.878	0.061	0.94	-1.5	0.39	1.40
36	bstw18_c	64.97	-0.695	0.047	0.97	-2.1	0.40	0.93
37	bevo01_c	77.80	-1.352	0.051	0.99	-0.5	0.32	0.79
38	bfkt09_c	55.55	-0.230	0.044	1.08	6.7	0.21	0.24
39	bfkt17_c	44.23	0.270	0.044	0.98	-1.9	0.39	0.85
40	bfkt19_c	75.86	-1.230	0.050	0.93	-2.9	0.44	1.38
41	bgen05_c	41.17	0.407	0.044	0.96	-3.2	0.42	0.96
42	bgen11_c	26.16	1.147	0.049	1.01	0.3	0.29	0.61
43	binf07_c	85.45	-1.895	0.060	0.99	-0.3	0.29	0.82
44	binf19_c	50.86	-0.021	0.044	0.98	-1.7	0.38	0.74
45	boek16_c	44.07	0.277	0.044	1.02	1.5	0.32	0.56
46	boek20_c	57.02	-0.292	0.044	0.98	-1.2	0.37	0.76
47	bstw13_c	70.00	-0.912	0.047	1.00	-0.0	0.33	0.74
48	bstw16_c	75.72	-1.225	0.050	0.98	-0.8	0.33	0.84
49	bevo02_c	81.75	-1.622	0.056	0.99	-0.2	0.29	0.72

	Item	Percentage cor-rect	Difficulty/ loca-tion parameter	*SE* (difficulty/ loca-tion parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination-2 PL
50	bfkt03_c	91.78	-2.575	0.077	0.99	-0.1	0.23	0.89
51	bfkt21_c	36.12	0.713	0.054	0.95	-2.6	0.42	0.94
52	bgen21_c	n.a.	0.910	0.079	1.00	0.1	0.22	0.31
53	binf05_c	59.83	-0.433	0.046	1.06	4.0	0.25	0.38
54	binf20_c	54.65	-0.199	0.046	1.02	1.6	0.32	0.56
55	boek02_c	77.45	-1.337	0.052	0.98	-0.6	0.34	0.86
56	boek12_c	37.53	0.569	0.046	1.04	2.8	0.27	0.44
57	boek13_c	37.22	0.587	0.046	1.01	0.5	0.33	0.61
58	bstw06_c	47.50	0.121	0.045	0.93	-5.9	0.48	1.17
59	bstw09_c	88.80	-2.223	0.067	0.96	-0.8	0.31	1.24
60	bstw11_c	34.21	0.733	0.048	1.00	0.2	0.33	0.62

Table 5

Step Parameters (and Standard Errors) of the Polytomous Item

Item	Step 1 (*SE*)	Step 2 (*SE*)
binf21_c	-0.281 (0.037)	0.281
bstw21_c	-0.022 (0.280)	0.022
bgen21_c	-1.266 (0.048)	1.266

5.2.2 Person parameters

The person parameters were estimated as WLEs (Pohl & Carstensen, 2012). WLEs will be provided in the next release of the SUF. A description of the data in the SUF can be found in Section 7. An overview of how to work with competence data is presented by Pohl and Carstensen (2012).

5.2.3 Test targeting and reliability

Test targeting focuses on how well item difficulties and person abilities are matched; this is an important criterion for evaluating the appropriateness of the test for the target group. In Figure 5, the item difficulties and person abilities are plotted on the same scale. The items covered the lower and medium part of the ability distribution very well. Hence, the test can measure person abilities in the low-ability and medium regions precisely, whereas high person abilities are measured with larger standard errors of measurement.

The mean of the ability distribution was constrained to be zero, and its variance was estimated to be 0.44, indicating a reasonable differentiation between the subjects. The reliability of the test (EAP/PV reliability = .73, WLE reliability = .73) was acceptable. This might be related to the suboptimal discrimination of test items.

Scale (in logits)	Person ability	Item difficulty
		24
2		
		30
	X	
	X	
	X	
	X	8
	XX	7
	XX	
	XXX	42
1	XX	
	XXXX	52
	XXXX	
	XXXX	51 60
	XXXXX	
	XXXXX	3 18 56 57
	XXXXXX	28 29 15
	XXXXX	41
	XXXXXXX	2 39 45
	XXXXXXXX	13
	XXXXXXXXX	34 58
	XXXXXXX	5 10
0	XXXXXXXX	25 44
	XXXXXXX	27 20
	XXXXXXXXX	14 38 54
	XXXXXXX	46
	XXXXXXX	1 26 31 53
	XXXXXX	
	XXXXXX	32
	XXXXX	36 23
	XXXXX	4
	XXXX	12
	XXXX	21 47
-1	XXXX	17
	XXX	19
	X	16 40 48
	X	22
	X	37 55
	X	
		11 33 49
		35 43
-2		6
		59
		9 50

Figure 5. Test targeting. The distribution of person abilities in the sample is depicted on the left-hand side, with each 'X' representing 28.7 cases. The item difficulties (or location parameters) are depicted on the right-hand side. Each number represents one item with a corresponding position in the test, cf. Table 3. Item difficulties of the steps of each location parameter are displayed using the suffix .1 for the first and .2 for the second step.

5.3 Quality of the Test

5.3.1 Item fit

Altogether, the item fit could be considered moderate, with values of the WMNSQ ranging from 0.93 (items bfkt19_c and bstw06_c) to 1.09 (item bfkt18_c), cf. column 5 of Table 4. Point-biserial correlations between the item scores and the total scores ranged from 0.13 (item bgen18_c) to 0.48 (item bstw06_c). Discriminations estimated in the generalized partial credit model with the TAM package in R ranged from 0.18 (item bfkt18_c) to 1.40 (item bstw04_c), cf. Table 4, column 8.

5.3.2 Differential item functioning

Differential item functioning (DIF) was used to evaluate test fairness for several subgroups (i. e., measurement invariance). For this purpose, DIF was examined for the variables gender, immigration background, books, and wave (see Pohl & Carstensen, 2012, for a description of these variables). Table 6 provides a summary of the results of the DIF analyses. According to Pohl and Carstensen (2012), absolute difficulty differences greater than 1 logit can be considered to show very strong DIF. For the current test, no item exceeded this threshold.

The table depicts the differences in the estimated item difficulties between the respective groups. "Male vs. Female", for example, indicates the difference in difficulty $ß_{male}$ - $ß_{female}$. A positive value indicates a higher difficulty for males, whereas a negative value indicates a lower difficulty for males as opposed to females.

Gender: On average, male participants had a considerably higher biological competence (main effect = -0.136 logits, Cohen's d = -0.205). [4] Five items (bfkt18_c, binf10_c, boek17_c, binf11_c, bstw21_c) showed a DIF greater than 0.6 logits.

Immigration background: On average, participants without an immigration background had a higher biological competence (main effect = 0.160 logits, Cohen's d = 0.241). No item showed a DIF greater than 0.6 logits.

Wave: On average, participants in the three waves differed only slightly in their biological competence (1 vs 2: main effect = -0.037, Cohen's d = -0.056; 1 vs 3: main effect = -0.081, Cohen's d = -0.122; 2 vs 3: main effect = -0.044, Cohen's d = -0.066). One item (bstw21_c) showed a DIF greater than 0.6 logits.

Books: On average, participants with many books at home performed better on the biological competence test (0-200 vs 201-500: main effect = 0.166, Cohen's d = 0.250; 0-200 vs > 500: main effect = 0.320, Cohen's d = 0.482; 201-500 vs > 500: main effect = 0.154, Cohen's d = 0.232). One item (bstw21_c) showed a DIF greater than 0.6 logits.

---

[4] The variance of the Partial-Credit Model was used to estimate the effect size.

Table 6

Differential Item Functioning

	Item	Gender	Immigration background	Wave			Books		
		male vs female	without vs with	1 vs 2	1 vs 3	2 vs 3	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
1	bevo16_c	0.094	0.092	-0.139	-0.080	0.059	0.094	0.029	-0.065
2	bfkt06_c	-0.350	0.024	0.015	-0.003	-0.018	0.058	0.046	-0.012
3	bfkt16_c	0.322	0.080	-0.220	-0.341	-0.121	-0.038	-0.201	-0.163
4	bfkt18_c	0.688	-0.060	-0.064	-0.254	-0.190	-0.015	-0.339	-0.324
5	bgen08_c	0.060	-0.234	-0.047	-0.055	-0.008	0.071	0.051	-0.020
6	binf10_c	0.800	-0.072	0.126	0.030	-0.096	-0.125	-0.240	-0.115
7	binf21_c	0.060	0.002	-0.066	-0.057	0.009	0.044	0.214	0.170
8	boek17_c	-0.694	0.088	0.148	0.070	-0.078	0.054	0.372	0.318
9	boek19_c	-0.502	0.204	-0.191	-0.022	0.169	0.108	0.249	0.141
10	bstw07_c	-0.144	0.118	0.061	0.095	0.034	-0.003	0.013	0.016
11	bstw12_c	-0.208	-0.024	-0.150	-0.293	-0.143	0.018	0.024	0.006
12	bstw19_c	0.118	-0.174	0.166	0.413	0.247	-0.159	-0.087	0.072
13	bevo20_c	-0.218	-0.070	0.213	0.018	-0.195	-0.124	0.034	0.158
14	bfkt10_c	-0.072	0.000	-0.106	-0.161	-0.055	-0.055	-0.023	0.032

	Item	Gender	Immigration background	Wave			Books		
		male vs female	without vs with	1 vs 2	1 vs 3	2 vs 3	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
15	bfkt13_c	-0.018	-0.178	-0.101	-0.163	-0.062	-0.106	0.073	0.179
16	bfkt14_c	0.200	-0.122	-0.052	0.159	0.211	-0.032	-0.316	-0.284
17	bgen14_c	0.004	-0.220	0.016	0.011	-0.005	0.044	-0.092	-0.136
18	binf09_c	-0.284	0.164	-0.117	0.054	0.171	-0.078	0.003	0.081
19	binf11_c	0.656	0.218	0.036	-0.366	-0.402	0.556	0.371	-0.185
20	boek04_c	-0.052	-0.010	-0.328	-0.491	-0.163	-0.160	-0.114	0.046
21	boek10_c	-0.262	-0.010	0.347	0.325	-0.022	0.085	-0.063	-0.148
22	bstw08_c	0.290	0.144	-0.236	0.068	0.304	-0.004	-0.141	-0.137
23	bstw17_c	0.378	0.034	0.222	0.300	0.078	-0.045	-0.182	-0.137
24	bstw21_c	-1.028	0.230	0.470	-0.734	-1.204	-0.802	-1.061	-0.259
25	bevo07_c	0.102	0.136	0.153	0.276	0.123	0.027	0.006	-0.021
26	bfkt02_c	-0.082	-0.146	0.370	0.316	-0.054	0.062	-0.013	-0.075
27	bfkt04_c	-0.252	-0.386	0.041	0.041	0.000	0.065	0.088	0.023
28	bfkt07_c	0.506	0.058	0.176	0.329	0.153	0.319	0.098	-0.221
29	bgen04_c	0.206	-0.200	0.212	0.358	0.146	-0.433	-0.455	-0.022
30	bgen18_c	0.102	0.136	0.112	0.224	0.112	-0.081	-0.081	0.000

	Item	Gender	Immigration background	Wave			Books		
		male vs female	without vs with	1 vs 2	1 vs 3	2 vs 3	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
31	binf08_c	0.490	-0.066	-0.115	-0.130	-0.015	-0.094	-0.245	-0.151
32	binf17_c	0.296	0.190	-0.052	0.013	0.065	-0.100	-0.164	-0.064
33	boek03_c	-0.072	0.118	0.045	-0.069	-0.114	0.206	0.592	0.386
34	boek09_c	-0.088	0.050	-0.003	-0.066	-0.063	0.177	0.302	0.125
35	bstw04_c	-0.154	0.106	-0.071	-0.400	-0.329	0.204	0.558	0.354
36	bstw18_c	-0.068	-0.130	-0.148	-0.242	-0.094	-0.070	-0.140	-0.070
37	bevo01_c	0.068	0.148	-0.228	-0.324	-0.096	0.195	0.360	0.165
38	bfkt09_c	0.344	-0.026	0.063	0.039	-0.024	-0.083	-0.355	-0.272
39	bfkt17_c	0.014	-0.114	0.013	0.089	0.076	0.001	-0.099	-0.100
40	bfkt19_c	-0.580	0.078	-0.326	-0.466	-0.140	-0.106	-0.066	0.040
41	bgen05_c	0.306	-0.014	0.134	0.280	0.146	0.088	0.062	-0.026
42	bgen11_c	-0.082	0.030	-0.152	0.017	0.169	0.044	0.286	0.242
43	binf07_c	0.342	0.278	0.042	-0.105	-0.147	0.156	0.101	-0.055
44	binf19_c	-0.212	0.054	0.194	0.367	0.173	0.307	0.142	-0.165
45	boek16_c	0.182	0.094	-0.185	-0.049	0.136	-0.068	0.071	0.139
46	boek20_c	-0.486	-0.014	0.165	0.150	-0.015	-0.179	0.056	0.235

	Item	Gender	Immigration background	Wave			Books		
		male vs female	without vs with	1 vs 2	1 vs 3	2 vs 3	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
47	bstw13_c	-0.398	-0.078	-0.280	-0.318	-0.038	-0.016	-0.137	-0.121
48	bstw16_c	-0.364	0.094	-0.096	-0.168	-0.072	-0.003	-0.224	-0.221
49	bevo02_c	-0.032	0.044	-0.091	-0.166	-0.075	0.122	-0.001	-0.123
50	bfkt03_c	-0.094	-0.124	0.232	0.188	-0.044	-0.193	-0.138	0.055
51	bfkt21_c	-0.156	0.310	0.071	0.219	0.148	0.131	0.184	0.053
52	bgen21_c	0.262	0.054	0.183	0.335	0.152	-0.103	0.130	0.233
53	binf05_c	-0.322	0.252	0.054	0.141	0.087	-0.160	-0.236	-0.076
54	binf20_c	0.096	0.090	0.120	0.219	0.099	-0.097	-0.100	-0.003
55	boek02_c	-0.154	-0.078	0.088	0.089	0.001	0.096	0.357	0.261
56	boek12_c	-0.348	-0.300	0.006	-0.126	-0.132	-0.368	-0.100	0.268
57	boek13_c	-0.310	-0.172	-0.057	0.159	0.216	0.004	0.464	0.460
58	bstw06_c	-0.114	0.250	-0.036	-0.153	-0.117	0.191	0.217	0.026
59	bstw09_c	0.196	-0.146	0.264	0.051	-0.213	0.041	0.052	0.011
60	bstw11_c	-0.358	-0.084	0.003	-0.009	-0.012	-0.014	-0.010	0.004
	main effect	-0.136	0.160	-0.037	-0.081	-0.044	0.166	0.320	0.154

In Table 7, the models with DIF are compared with those that included only the main effect of the respective variable. Regarding Akaike's (1974) information criterion (AIC), the more parsimonious models including only main effects were preferred over the ones containing the variable immigration background. The Bayesian information criterion (BIC; Schwarz, 1978) takes into account the number of estimated parameters and, thus, prevents the overparameterization of models. Using BIC, the more complex model including DIF was preferred only for the variable gender.

Table 7

Comparison of Models With and Without DIF

DIF variable	Model	Number of parameters	AIC	BIC
Gender	main effect	65	195,622.91	195,732.67
	DIF	128	194,930.78	195,146.92
Immigration Background	main effect	65	194,296.28	194,406.04
	DIF	127[5]	194,299.44	194,513.89
Wave	main effect	66	196,435.23	196,546.68
	DIF	191[6]	196,410.21	196,732.73
Books	main effect	66	195,446.45	195,557.90
	DIF	192	195,369.12	195,693.33

5.3.3 Rasch homogeneity

One essential assumption of the partial credit model is Rasch homogeneity. Rasch homogeneity implies that all item-discrimination parameters are equal. In order to test this assumption, a generalized partial credit model (2PL) was specified (see Table 4). In this model, discrimination parameters are freely estimated and not fixed to 1. The estimated discriminations differed across the items (see Table 4), ranging from 0.18 (item bfkt18_c) to 1.40 (item bstw04_c). Despite the empirical preference for the generalized partial credit model (AIC = 195,458, BIC = 196,257, number of parameters = 123) model, the partial credit model (AIC = 196,438, BIC = 196,853, number of parameters = 64) more adequately matched with the theoretical conceptions underlying the construction of the test (see Pohl & Carstensen, 2012, 2013 for a discussion of this issue). For this reason, the 1PL partial credit model was chosen as the scaling model.

6. Discussion

Descriptions and analyses presented in the previous sections were aimed at documenting the quality of the biological competence test used in the additional study Baden-Wuerttemberg. The occurrence of different kinds of missing responses was evaluated, and item as well as test quality was examined. Furthermore, measurement invariance was examined with regard to various grouping variables. The item fit statistics provided evidence of items with good fit that

---

[5] No person with immigration background completely solved this item, therefore the degrees of freedom (df) reduced by 1 df.
[6] No person in wave 3 completely solved this item, therefore the degrees of freedom (df) reduced by 1 df.

were measurement invariant across these subgroups. The test was found to be reasonably reliable. As shown, ability estimates for participants with low performance were found to be precise but less precise for medium- and high-performing participants.

7. Data in the Scientific Use File

The data in the Scientific Use File contain 60 items, of which 57 are scored as dichotomous variables with 0 (*incorrect response*) or 1 (*correct response*). Overall, 3 items were polytomous and in accordance with Pohl & Carstensen (2012) scored either 0 (*incorrect response*), 0.5 (*partially correct response*), or 1 (*correct response*). MC items are marked with a '_c' at the end of the variable name. Appendix A provides the syntax that was used to generate the person estimates with the ConQuest 4.2 software (Wu, Adams, & Wilson, 1997). Appendix B provides an alternative syntax for use with the TAM package (Kiefer, Robitzsch, & Wu, 2015) in the software R (R Core Team, 2015).

Manifest biological competence scores are provided in the form of WLEs (b_sc1) along with their corresponding standard errors (b_sc2). As described in Section 5, these person estimates were derived from the joint scaling of all three waves of the study. For persons who did not take the biological competence test, no WLE was estimated. WLEs were estimated for all items delivered in the Scientific Use File. No item showed negative discrimination in the 2PL model, so we did not have to exclude any items from further analyses. In order to allow the users to estimate their own WLEs by considering different item selection standards, all test items are delivered in the Scientific Use File. For researchers interested in analysis that require one of the variables that showed DIF > 0.6 logits, we emphasize that models should be considered on the basis partial measurement invariance (e.g. Byrne, Shavelson & Muthén, 1989).

We recommend the use of plausible values to investigate latent relationships between competence scores and other variables. Users interested in examining latent relationships may either include the measurement model in their analyses or estimate plausible values themselves. A description of these approaches can be found in Pohl and Carstensen (2012).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–722.

Biologie heute entdecken. (2004*). Allgemeine Ausgabe für SII.* Braunschweig: Schroedel.

Biologie Oberstufe. (2001). *Gesamtband.* Berlin: Cornelsen.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the Equivalence of Factor Covariance and Mean Structure: The Issue of Partial Measurement Invariance*. Psychological Bulletin, 105*, 456-466.

Duchhardt, C. (2015). *NEPS Technical Report for Mathematics: Scaling results for the additional study Baden-Wuerttemberg* (NEPS Working Paper No. 59). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Eberle, F., Gehrer, K., Jaggi, B., Kottonau, J., Oepke, M. & Pflüger, M. (2008). *Evaluation der Maturitätsreform 1995. Schlussbericht zur Phase II.* Bern: Staatssekretariat für Bildung und Forschung SBF.

Hardt, K., Pohl, S., Haberkorn, K., & Wiegand, E. (2013). *NEPS Technical Report for Reading—Scaling results of Starting Cohort 6 for adults in main study 2010/11* (NEPS Working Paper No. 25). Bamberg: University of Bamberg, National Educational Panel Study.

Hübner, N., Rieger, S., & Wagner, W. (2016). *NEPS Technical Report for English Reading – Scaling Results of the Additional Study Baden-Württemberg* (NEPS Working Paper No. X). Bamberg: Leibniz-Institute for Educational Trajectories, National Educational Panel Study.

Jordan, A.-K., & Duchhardt, C. (2013). *NEPS Technical Report for Mathematics—Scaling results of Starting Cohort 6–Adults* (NEPS Working Paper No. 32). Bamberg: University of Bamberg, National Educational Panel Study.

Kiefer, T., Robitzsch, A., & Wu, M. (2015). *TAM: Test Analysis Modules (R package version 1.4-1)* [Computer software]. Retrieved from http://CRAN.R-project.org/package=TAM

Koller, I., Haberkorn, K., & Rohm, T. (2014). *NEPS Technical Report for Reading: Scaling results of Starting Cohort 6 for adults in main study 2012* (NEPS Working Paper No. 48). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Linder Biologie (2005*). Gesamtband SII,* 22. Auflage. Braunschweig: Schroedel.

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

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14,* 59-71.

Natura (2006). *Grundlagen der Biologie für Schweizerische Maturitätsschulen*. Zug: Klett und Balmer.

NEPS (2011). *G8-Reform in Baden-Württemberg, Haupterhebung 2010/11 (A72), Schüler/innen, Klasse 13. Informationen zum Kompetenztest.* Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/BW/2-0-0/C_A72_de.pdf.

NEPS (2012). *G8-Reform in Baden-Württemberg, Haupterhebung 2011/12 (A73), Schüler/innen, Klasse 12/13. Informationen zum Kompetenztest.* Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/BW/2-0-0/C_A73_de.pdf.

Pohl, S., & Carstensen, C. H. (2012). *NEPS Technical report – Scaling the data of the competence tests* (NEPS Working Paper No. 14). Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Pohl, S., Haberkorn, K., Hardt, K., & Wiegand, E. (2012). *NEPS Technical Report for Reading– Scaling results of Starting Cohort 3 in fifth grade* (NEPS Working Paper No. 15). Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

R Core Team (2015). R: *A language and environment for statistical computing*. R Foundationfor Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2004). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie* (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 05.02.2004).

Wagner, W., Kramer, J., Trautwein, U., Lüdtke, O., Nagy, G., Jonkmann, K., Maaz, K., Meixner, S., & Schilling, J. (2011). 15 Upper secondary education in academic school tracks and the transition from school to postsecondary education and the job market. *Zeitschrift für Erziehungswissenschaft*, *14*, 233-249.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54,* 427–450.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software [computer program]*. Camberwell, Vic.: Australian Council for Educational Research.

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF). *Ottawa: National Defense Headquarters*.

Appendix

Appendix A: ConQuest Syntax for generating WLE estimates in the Additional Study Baden-Wuerttemberg

title Additional Study Baden-Wuerttemberg, biological competence, Waves 1-3, Partial credit model;

datafile filename.dat;

format pid 1-7 responses 12-70;

labels << labels.nam;

codes 0,1,2;

score (0,1) (0,1) !item(1-6,8-35,37-51,53-60);

score (0,1,2) (0,.5,1) !item(7,36,52);

model item + item*step;

set constraint=cases;

estimate ! stderr=empirical;

itanal ! form=long >> filename.itn;

export parameters >> filename.prm;

show cases ! estimates=wle >> filename.wle;

show ! estimates=latent, tables=1:2:3:4:5 >> filename.shw;

Appendix B: TAM Syntax for generating WLE estimates in the Additional Study Baden-Wuerttemberg

```
setwd ("Your/Working/Directory")

data <- # data read

items <- # column positions of the biological competence items in the SUF

library (TAM)

# Generate Design Matrices

Des <- designMatrices(modeltype="PCM", resp = data[,items]])

B <- Des$B

# Score the three polytomous items according to NEPS conventions

B[,3,][7] <- 1

B[,2,][7] <- 0.5

B[,3,][36] <- 1

B[,2,][36] <- 0.5

B[,3,][52] <- 1

B[,2,][52] <- 0.5

B

# Compute PCM

PCM <- tam.mml (data[,items],irtmodel="PCM",B=B,pid=data$id)

pcm.mod.bio$item

summary (PCM)

# Generate WLE estimates

PCM.wle <- tam.wle(PCM)

WLE <- PCM.wle$theta

WLE.SE <- PCM.wle$error
```

Table 8

Item Parameters of the Biological Competence Test – Wave 1

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
1	bevo16_c	61.53	-0.520	0.063	1.00	0.1	0.34
2	bfkt06_c	43.05	0.308	0.062	1.02	1.2	0.32
3	bfkt16_c	41.46	0.379	0.064	1.08	4.0	0.21
4	bfkt18_c	69.37	-0.896	0.066	1.08	3.1	0.18
5	bgen08_c	51.12	-0.056	0.062	1.05	2.8	0.27
6	binf10_c	85.94	-1.956	0.085	1.00	0.1	0.24
7	binf21_c	n.a.	1.230	0.094	0.95	-1.4	0.38
8	boek17_c	21.96	1.384	0.074	0.95	-1.4	0.38
9	boek19_c	94.06	-2.942	0.122	1.00	0.0	0.17
10	bstw07_c	48.21	0.078	0.062	0.97	-1.6	0.40
11	bstw12_c	84.57	-1.839	0.082	0.96	-0.7	0.34
12	bstw19_c	64.95	-0.678	0.064	0.98	-0.8	0.37
13	bevo20_c	41.63	0.327	0.095	1.00	0.2	0.34
14	bfkt10_c	56.75	-0.318	0.088	0.97	-1.2	0.40

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
15	bfkt13_c	42.31	0.316	0.089	1.00	0.0	0.33
16	bfkt14_c	74.96	-1.213	0.101	1.00	0.1	0.31
17	bgen14_c	71.48	-1.019	0.096	1.03	0.8	0.25
18	binf09_c	37.79	0.517	0.091	1.05	1.6	0.24
19	binf11_c	75.53	-1.250	0.100	0.99	-0.2	0.33
20	boek04_c	59.69	-0.446	0.091	1.02	0.7	0.30
21	boek10_c	66.28	-0.760	0.092	1.00	0.1	0.34
22	bstw08_c	78.23	-1.418	0.103	1.07	1.2	0.16
23	bstw17_c	62.07	-0.561	0.090	0.95	-1.9	0.44
24	bstw21_c	n.a.	5.997	0.997	0.97	-0.1	0.27
25	bevo07_c	49.21	0.025	0.091	1.02	0.8	0.33
26	bfkt02_c	53.11	-0.150	0.089	1.01	0.3	0.34
27	bfkt04_c	52.21	-0.112	0.089	1.02	0.9	0.32
28	bfkt07_c	35.27	0.652	0.092	0.96	-1.2	0.41
29	bgen04_c	35.68	0.630	0.092	1.05	1.5	0.25
30	bgen18_c	16.10	1.778	0.118	1.04	0.5	0.17
31	binf08_c	60.80	-0.503	0.091	0.98	-0.6	0.36

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
32	binf17_c	64.53	-0.674	0.092	1.01	0.2	0.33
33	boek03_c	82.07	-1.670	0.111	0.97	-0.4	0.34
34	boek09_c	47.37	0.094	0.089	1.02	0.9	0.31
35	bstw04_c	87.00	-2.062	0.126	0.96	-0.5	0.35
36	bstw18_c	68.55	-0.868	0.094	0.96	-1.0	0.40
37	bevo01_c	81.39	-1.585	0.105	1.00	0.0	0.28
38	bfkt09_c	55.52	-0.228	0.085	1.07	3.0	0.21
39	bfkt17_c	44.58	0.260	0.086	0.99	-0.6	0.38
40	bfkt19_c	80.91	-1.550	0.104	0.92	-1.3	0.44
41	bgen05_c	39.04	0.503	0.087	0.95	-2.0	0.44
42	bgen11_c	28.02	1.038	0.094	0.99	-0.3	0.34
43	binf07_c	86.02	-1.939	0.118	0.99	-0.1	0.28
44	binf19_c	47.56	0.125	0.086	0.99	-0.3	0.36
45	boek16_c	47.21	0.136	0.085	1.02	1.0	0.32
46	boek20_c	55.28	-0.213	0.086	1.00	0.1	0.33
47	bstw13_c	74.77	-1.168	0.095	1.00	0.0	0.32
48	bstw16_c	77.88	-1.353	0.099	0.96	-0.8	0.39

	Item	Percentage cor-rect	Difficulty/ loca-tion parameter	*SE* (difficulty/ loca-tion parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
49	bevo02_c	83.64	-1.746	0.111	0.98	-0.2	0.30
50	bfkt03_c	90.98	-2.455	0.142	1.00	0.1	0.21
51	bfkt21_c	35.33	0.760	0.104	0.94	-1.6	0.44
52	bgen21_c	n.a.	1.047	0.163	1.01	0.1	0.19
53	binf05_c	59.68	-0.410	0.088	1.07	2.7	0.23
54	binf20_c	53.26	-0.125	0.089	1.03	1.4	0.29
55	boek02_c	77.28	-1.310	0.101	0.98	-0.3	0.34
56	boek12_c	39.19	0.505	0.087	1.06	2.1	0.24
57	boek13_c	38.06	0.559	0.089	1.01	0.3	0.33
58	bstw06_c	49.92	0.029	0.087	0.93	-3.1	0.48
59	bstw09_c	87.99	-2.124	0.126	0.97	-0.3	0.31
60	bstw11_c	35.34	0.694	0.091	0.99	-0.4	0.35

Table 9

Differential Item Functioning – Wave 1

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
1	bevo16_c	0.030	0.188	0.039	0.108	0.069
2	bfkt06_c	-0.380	0.150	0.222	-0.020	-0.242
3	bfkt16_c	0.398	0.022	0.186	-0.066	-0.252
4	bfkt18_c	0.788	-0.202	-0.184	-0.434	-0.250
5	bgen08_c	0.156	-0.234	0.098	0.058	-0.040
6	binf10_c	0.856	-0.024	0.199	-0.079	-0.278
7	binf21_c	-0.038	-0.134	-0.202	0.070	0.272
8	boek17_c	-0.576	0.234	-0.099	0.379	0.478
9	boek19_c	-0.994	0.186	-0.300	-0.060	0.240
10	bstw07_c	-0.270	-0.026	-0.183	0.086	0.269
11	bstw12_c	0.080	-0.126	0.153	0.198	0.045
12	bstw19_c	0.202	-0.008	-0.011	-0.111	-0.100
13	bevo20_c	-0.030	-0.100	-0.094	0.151	0.245
14	bfkt10_c	0.048	-0.178	-0.089	-0.121	-0.032

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
15	bfkt13_c	-0.178	-0.174	-0.464	0.137	0.601
16	bfkt14_c	-0.094	-0.104	-0.025	-0.421	-0.396
17	bgen14_c	0.248	-0.390	0.089	0.241	0.152
18	binf09_c	-0.380	0.258	-0.419	-0.064	0.355
19	binf11_c	0.728	0.082	0.342	0.537	0.195
20	boek04_c	0.084	0.110	0.038	-0.231	-0.269
21	boek10_c	-0.540	0.112	0.052	-0.325	-0.377
22	bstw08_c	0.314	0.288	0.277	0.170	-0.107
23	bstw17_c	0.566	-0.232	-0.012	-0.116	-0.104
24	bstw21_c	-0.672	-0.662	-0.162	-0.528	-0.366
25	bevo07_c	0.472	0.344	0.027	-0.168	-0.195
26	bfkt02_c	-0.304	0.056	0.244	-0.070	-0.314
27	bfkt04_c	-0.238	-0.314	0.298	0.242	-0.056
28	bfkt07_c	0.388	0.098	0.069	-0.183	-0.252
29	bgen04_c	0.408	-0.374	-0.474	-0.162	0.312
30	bgen18_c	0.320	-0.188	-0.051	0.069	0.120

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
31	binf08_c	0.382	-0.250	-0.037	-0.221	-0.184
32	binf17_c	0.406	0.276	-0.011	-0.664	-0.653
33	boek03_c	-0.394	0.212	-0.088	0.436	0.524
34	boek09_c	-0.100	-0.132	0.012	-0.099	-0.111
35	bstw04_c	-0.022	-0.152	0.632	0.339	-0.293
36	bstw18_c	-0.046	-0.362	-0.195	-0.510	-0.315
37	bevo01_c	0.380	-0.022	0.008	0.166	0.158
38	bfkt09_c	0.322	0.180	-0.196	-0.308	-0.112
39	bfkt17_c	-0.136	-0.302	-0.092	-0.265	-0.173
40	bfkt19_c	-0.738	-0.004	-0.220	-0.644	-0.424
41	bgen05_c	0.278	0.060	-0.084	-0.148	-0.064
42	bgen11_c	-0.162	-0.142	-0.099	0.072	0.171
43	binf07_c	0.290	0.414	-0.140	0.166	0.306
44	binf19_c	-0.332	0.116	-0.028	-0.159	-0.131
45	boek16_c	0.100	0.018	-0.088	0.176	0.264
46	boek20_c	-0.324	-0.076	-0.545	0.011	0.556

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
47	bstw13_c	-0.336	0.094	-0.011	-0.212	-0.201
48	bstw16_c	-0.384	0.154	0.089	-0.176	-0.265
49	bevo02_c	-0.088	-0.282	0.128	-0.250	-0.378
50	bfkt03_c	-0.194	-0.458	-0.509	-0.366	0.143
51	bfkt21_c	-0.176	0.192	-0.038	-0.004	0.034
52	bgen21_c	0.264	0.248	-0.143	0.857	1.000
53	binf05_c	-0.422	0.346	-0.144	-0.222	-0.078
54	binf20_c	0.288	0.012	-0.123	-0.108	0.015
55	boek02_c	-0.082	0.106	0.563	0.550	-0.013
56	boek12_c	-0.326	0.020	-0.530	-0.240	0.290
57	boek13_c	-0.640	-0.092	0.071	0.740	0.669
58	bstw06_c	-0.374	0.610	0.180	-0.108	-0.288
59	bstw09_c	-0.230	0.024	0.141	-0.073	-0.214
60	bstw11_c	-0.288	-0.054	-0.061	0.362	0.423
	main effect	-0.182	0.196	0.092	0.298	0.206

Appendix C2: Item Parameters and Differential Item Functioning for Wave 2 from the Additional Study Baden-Wuerttemberg only

Table 10

Item Parameters of the Biological Competence Test – Wave 2

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
1	bevo16_c	57.73	-0.344	0.046	1.00	-0.4	0.35
2	bfkt06_c	42.43	0.330	0.046	1.00	-0.3	0.35
3	bfkt16_c	35.78	0.635	0.048	1.04	2.4	0.26
4	bfkt18_c	67.35	-0.793	0.048	1.08	4.3	0.18
5	bgen08_c	49.24	0.028	0.046	1.04	3.1	0.28
6	binf10_c	86.94	-2.042	0.064	1.02	0.5	0.19
7	binf21_c	n.a.	1.331	0.072	0.95	-2.2	0.39
8	boek17_c	23.68	1.271	0.052	0.96	-1.6	0.38
9	boek19_c	92.70	-2.712	0.082	0.99	-0.1	0.20
10	bstw07_c	48.76	0.054	0.045	0.99	-1.1	0.37
11	bstw12_c	82.14	-1.650	0.057	0.96	-1.2	0.36
12	bstw19_c	67.60	-0.805	0.048	1.00	0.2	0.34
13	bevo20_c	46.44	0.150	0.066	1.00	0.1	0.35
14	bfkt10_c	53.71	-0.175	0.063	0.98	-1.4	0.38

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
15	bfkt13_c	39.63	0.453	0.066	0.98	-1.2	0.38
16	bfkt14_c	73.62	-1.123	0.072	1.06	1.7	0.21
17	bgen14_c	71.21	-0.997	0.069	1.03	1.1	0.27
18	binf09_c	34.84	0.669	0.067	1.03	1.4	0.26
19	binf11_c	75.67	-1.248	0.073	0.99	-0.3	0.32
20	boek04_c	51.53	-0.080	0.066	1.03	1.9	0.27
21	boek10_c	72.62	-1.068	0.070	1.01	0.3	0.30
22	bstw08_c	73.94	-1.144	0.071	1.02	0.6	0.27
23	bstw17_c	66.23	-0.745	0.067	0.96	-1.5	0.40
24	bstw21_c	n.a.	5.559	0.580	1.00	0.0	0.15
25	bevo07_c	51.91	-0.091	0.066	1.03	1.6	0.31
26	bfkt02_c	60.48	-0.482	0.066	1.02	0.8	0.31
27	bfkt04_c	52.12	-0.115	0.065	1.03	1.5	0.29
28	bfkt07_c	38.04	0.513	0.066	0.96	-1.8	0.40
29	bgen04_c	39.30	0.455	0.066	1.05	2.2	0.25
30	bgen18_c	16.85	1.700	0.085	1.07	1.4	0.10
31	binf08_c	57.53	-0.350	0.066	0.98	-0.8	0.37

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
32	binf17_c	62.66	-0.583	0.067	1.00	-0.0	0.34
33	boek03_c	82.26	-1.676	0.081	1.00	-0.1	0.28
34	boek09_c	46.33	0.135	0.066	1.00	0.2	0.35
35	bstw04_c	85.74	-1.951	0.089	0.95	-0.9	0.36
36	bstw18_c	64.72	-0.682	0.067	0.97	-1.5	0.40
37	bevo01_c	77.37	-1.319	0.072	0.99	-0.3	0.33
38	bfkt09_c	56.08	-0.254	0.062	1.09	5.0	0.18
39	bfkt17_c	43.86	0.284	0.063	0.98	-0.9	0.37
40	bfkt19_c	75.17	-1.186	0.071	0.94	-1.7	0.42
41	bgen05_c	41.16	0.405	0.063	0.96	-1.9	0.40
42	bgen11_c	24.77	1.226	0.071	1.02	0.7	0.26
43	binf07_c	86.05	-1.943	0.086	0.98	-0.4	0.29
44	binf19_c	51.04	-0.031	0.062	0.97	-1.8	0.40
45	boek16_c	42.30	0.357	0.063	1.03	1.4	0.30
46	boek20_c	58.13	-0.340	0.063	0.97	-1.4	0.39
47	bstw13_c	68.85	-0.850	0.066	1.02	0.6	0.30
48	bstw16_c	75.67	-1.218	0.071	1.00	-0.1	0.32

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
49	bevo02_c	81.77	-1.616	0.079	1.00	-0.1	0.28
50	bfkt03_c	92.35	-2.648	0.113	0.98	-0.2	0.23
51	bfkt21_c	35.79	0.725	0.078	0.96	-1.5	0.41
52	bgen21_c	n.a.	0.900	0.110	1.00	0.1	0.21
53	binf05_c	59.72	-0.427	0.065	1.03	1.4	0.29
54	binf20_c	54.95	-0.207	0.065	1.02	1.1	0.31
55	boek02_c	77.93	-1.359	0.075	0.97	-0.7	0.35
56	boek12_c	38.28	0.534	0.065	1.04	1.7	0.28
57	boek13_c	35.80	0.652	0.066	1.01	0.5	0.32
58	bstw06_c	47.95	0.102	0.064	0.94	-3.8	0.47
59	bstw09_c	90.02	-2.348	0.101	0.96	-0.5	0.29
60	bstw11_c	34.26	0.727	0.068	1.00	0.1	0.33

Table 11

Differential Item Functioning – Wave 2

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
1	bevo16_c	0.138	-0.034	0.119	0.040	-0.079
2	bfkt06_c	-0.370	-0.002	0.029	0.079	0.050
3	bfkt16_c	0.326	0.132	-0.071	-0.232	-0.161
4	bfkt18_c	0.712	0.046	0.108	-0.186	-0.294
5	bgen08_c	-0.026	-0.248	0.076	0.083	0.007
6	binf10_c	0.820	-0.118	-0.212	-0.301	-0.089
7	binf21_c	0.104	-0.010	0.149	0.232	0.083
8	boek17_c	-0.748	0.156	0.046	0.248	0.202
9	boek19_c	-0.284	0.156	0.260	0.146	-0.114
10	bstw07_c	-0.020	0.154	0.091	-0.064	-0.155
11	bstw12_c	-0.284	-0.016	0.002	-0.088	-0.090
12	bstw19_c	0.012	-0.236	-0.236	-0.073	0.163
13	bevo20_c	-0.292	-0.140	-0.206	-0.015	0.191
14	bfkt10_c	0.014	-0.084	-0.235	-0.005	0.230

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
15	bfkt13_c	0.040	-0.194	0.011	-0.118	-0.129
16	bfkt14_c	0.302	-0.120	-0.053	-0.421	-0.368
17	bgen14_c	0.090	-0.178	-0.005	-0.211	-0.206
18	binf09_c	-0.294	0.164	-0.156	0.018	0.174
19	binf11_c	0.580	0.114	0.537	0.264	-0.273
20	boek04_c	-0.094	-0.106	-0.162	-0.061	0.101
21	boek10_c	-0.298	-0.072	0.053	0.074	0.021
22	bstw08_c	0.266	0.238	-0.289	-0.242	0.047
23	bstw17_c	0.344	0.348	0.190	-0.226	-0.416
24	bstw21_c	-0.540	0.532	-0.934	-2.246	-1.312
25	bevo07_c	-0.060	-0.202	0.031	-0.043	-0.074
26	bfkt02_c	0.058	-0.302	-0.215	0.058	0.273
27	bfkt04_c	-0.384	-0.480	-0.078	-0.173	-0.095
28	bfkt07_c	0.502	-0.026	0.285	0.138	-0.147
29	bgen04_c	0.070	0.068	-0.483	-0.597	-0.114
30	bgen18_c	-0.228	0.276	0.069	-0.021	-0.090

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
31	binf08_c	0.636	0.016	-0.120	-0.327	-0.207
32	binf17_c	0.202	0.114	-0.207	-0.198	0.009
33	boek03_c	-0.026	0.076	0.135	0.570	0.435
34	boek09_c	-0.182	0.022	0.166	0.328	0.162
35	bstw04_c	-0.388	-0.016	0.104	0.608	0.504
36	bstw18_c	0.010	-0.010	-0.049	0.027	0.076
37	bevo01_c	-0.054	0.188	0.095	0.481	0.386
38	bfkt09_c	0.440	0.018	0.111	-0.312	-0.423
39	bfkt17_c	0.084	0.022	-0.086	0.031	0.117
40	bfkt19_c	-0.620	-0.016	-0.081	0.088	0.169
41	bgen05_c	0.288	0.034	0.216	0.264	0.048
42	bgen11_c	0.272	0.196	0.231	0.447	0.216
43	binf07_c	0.480	0.232	0.509	0.124	-0.385
44	binf19_c	-0.090	0.044	0.543	0.197	-0.346
45	boek16_c	0.230	0.130	-0.046	0.097	0.143
46	boek20_c	-0.466	0.088	0.033	0.238	0.205

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
47	bstw13_c	-0.470	-0.012	0.001	-0.117	-0.118
48	bstw16_c	-0.396	0.048	0.130	-0.250	-0.380
49	bevo02_c	-0.048	0.216	-0.046	0.127	0.173
50	bfkt03_c	0.192	-0.116	-0.075	0.003	0.078
51	bfkt21_c	-0.160	0.474	0.033	0.142	0.109
52	bgen21_c	0.456	0.030	-0.184	-0.128	0.056
53	binf05_c	-0.372	0.284	-0.034	-0.092	-0.058
54	binf20_c	0.002	-0.008	-0.109	-0.117	-0.008
55	boek02_c	-0.258	-0.096	-0.252	0.111	0.363
56	boek12_c	-0.328	-0.422	-0.441	-0.018	0.423
57	boek13_c	-0.294	-0.110	0.026	0.205	0.179
58	bstw06_c	0.000	0.192	0.122	0.398	0.276
59	bstw09_c	0.302	-0.478	-0.295	0.051	0.346
60	bstw11_c	-0.416	-0.090	0.024	-0.212	-0.236
	main effect	-0.126	0.138	0.157	0.302	0.145

Table 12

Item Parameters of the Biological Competence Test – Wave 3

	Item	Percentage cor-rect	Difficulty/ loca-tion parameter	*SE* (difficulty/ loca-tion parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
1	bevo16_c	58.02	-0.361	0.065	0.98	-0.8	0.39
2	bfkt06_c	41.18	0.394	0.066	1.00	0.2	0.35
3	bfkt16_c	32.65	0.805	0.070	1.07	2.6	0.22
4	bfkt18_c	62.38	-0.564	0.066	1.12	5.4	0.14
5	bgen08_c	47.94	0.080	0.065	1.05	2.7	0.29
6	binf10_c	85.24	-1.913	0.086	1.04	0.7	0.21
7	binf21_c	n.a.	1.376	0.103	0.93	-2.0	0.42
8	boek17_c	21.86	1.402	0.076	0.94	-1.5	0.41
9	boek19_c	93.39	-2.851	0.120	1.00	-0.0	0.19
10	bstw07_c	48.56	0.065	0.065	1.00	-0.2	0.37
11	bstw12_c	79.28	-1.473	0.077	0.97	-0.7	0.37
12	bstw19_c	71.45	-1.014	0.070	1.00	0.1	0.33
13	bevo20_c	41.65	0.392	0.100	1.02	0.8	0.31
14	bfkt10_c	51.91	-0.077	0.089	0.99	-0.3	0.37

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
15	bfkt13_c	37.82	0.563	0.093	0.99	-0.4	0.36
16	bfkt14_c	76.78	-1.297	0.105	1.00	0.1	0.30
17	bgen14_c	70.58	-0.954	0.096	1.02	0.4	0.31
18	binf09_c	37.93	0.545	0.092	1.02	0.7	0.31
19	binf11_c	67.55	-0.807	0.094	0.97	-0.8	0.38
20	boek04_c	47.51	0.128	0.094	1.05	1.8	0.29
21	boek10_c	71.67	-1.008	0.097	1.04	0.9	0.27
22	bstw08_c	78.44	-1.412	0.104	0.99	-0.1	0.30
23	bstw17_c	67.22	-0.784	0.094	0.97	-0.7	0.40
24	bstw21_c	n.a.	6.827	0.647	0.97	0.0	0.24
25	bevo07_c	53.64	-0.172	0.095	1.03	1.0	0.33
26	bfkt02_c	58.25	-0.388	0.092	0.97	-1.0	0.41
27	bfkt04_c	51.14	-0.073	0.091	1.03	1.3	0.31
28	bfkt07_c	40.45	0.405	0.093	0.98	-0.6	0.37
29	bgen04_c	41.74	0.354	0.092	1.01	0.5	0.35
30	bgen18_c	18.07	1.642	0.116	1.07	1.1	0.16
31	binf08_c	56.23	-0.294	0.093	1.03	0.9	0.34

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
32	binf17_c	63.07	-0.610	0.095	0.98	-0.7	0.40
33	boek03_c	79.86	-1.529	0.109	0.98	-0.3	0.33
34	boek09_c	44.25	0.242	0.092	0.96	-1.7	0.44
35	bstw04_c	80.83	-1.589	0.112	0.90	-1.7	0.47
36	bstw18_c	61.72	-0.548	0.093	0.98	-0.6	0.40
37	bevo01_c	74.75	-1.186	0.100	0.99	-0.2	0.35
38	bfkt09_c	54.50	-0.186	0.089	1.08	2.9	0.25
39	bfkt17_c	44.57	0.254	0.090	0.97	-1.3	0.43
40	bfkt19_c	71.74	-1.007	0.097	0.92	-1.8	0.45
41	bgen05_c	43.50	0.307	0.090	0.94	-2.2	0.46
42	bgen11_c	26.92	1.109	0.099	0.99	-0.2	0.31
43	binf07_c	83.64	-1.761	0.116	1.00	0.1	0.31
44	binf19_c	54.05	-0.160	0.090	0.99	-0.4	0.38
45	boek16_c	44.26	0.267	0.090	1.01	0.4	0.34
46	boek20_c	56.67	-0.283	0.090	0.99	-0.2	0.37
47	bstw13_c	67.11	-0.773	0.094	0.98	-0.6	0.39
48	bstw16_c	73.48	-1.109	0.099	1.01	0.2	0.30

	Item	Percentage correct	Difficulty/ location parameter	*SE* (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score
49	bevo02_c	79.58	-1.507	0.110	1.01	0.3	0.27
50	bfkt03_c	91.49	-2.572	0.154	0.98	-0.1	0.23
51	bfkt21_c	37.72	0.628	0.112	0.94	-1.5	0.43
52	bgen21_c	n.a.	0.797	0.155	0.98	-0.3	0.27
53	binf05_c	60.21	-0.474	0.093	1.11	3.6	0.18
54	binf20_c	55.58	-0.265	0.093	1.01	0.5	0.36
55	boek02_c	76.64	-1.324	0.106	1.00	0.1	0.32
56	boek12_c	34.15	0.715	0.096	1.04	1.0	0.28
57	boek13_c	39.15	0.484	0.094	1.00	-0.1	0.37
58	bstw06_c	43.92	0.263	0.092	0.92	-3.0	0.49
59	bstw09_c	87.22	-2.103	0.130	0.94	-0.6	0.34
60	bstw11_c	32.84	0.788	0.099	1.01	0.2	0.31

Table 13

Differential Item Functioning – Wave 3

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
1	bevo16_c	0.080	0.242	0.097	-0.075	-0.172
2	bfkt06_c	-0.276	-0.058	-0.073	0.040	0.113
3	bfkt16_c	0.240	0.038	-0.255	-0.285	-0.030
4	bfkt18_c	0.550	-0.118	-0.090	-0.528	-0.438
5	bgen08_c	0.136	-0.206	0.026	-0.023	-0.049
6	binf10_c	0.714	-0.060	-0.304	-0.302	0.002
7	binf21_c	0.094	0.174	0.114	0.342	0.228
8	boek17_c	-0.714	-0.172	0.265	0.644	0.379
9	boek19_c	-0.572	0.340	0.174	0.783	0.609
10	bstw07_c	-0.256	0.206	0.001	0.077	0.076
11	bstw12_c	-0.326	0.052	-0.077	0.092	0.169
12	bstw19_c	0.250	-0.232	-0.164	-0.112	0.052
13	bevo20_c	-0.254	0.106	0.039	0.027	-0.012
14	bfkt10_c	-0.364	0.334	0.349	0.053	-0.296

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
15	bfkt13_c	0.036	-0.160	0.054	0.396	0.342
16	bfkt14_c	0.306	-0.134	0.002	0.004	0.002
17	bgen14_c	-0.402	-0.142	0.104	-0.182	-0.286
18	binf09_c	-0.170	0.088	0.413	0.046	-0.367
19	binf11_c	0.730	0.492	0.817	0.449	-0.368
20	boek04_c	-0.134	0.060	-0.359	-0.082	0.277
21	boek10_c	0.114	-0.050	0.202	-0.031	-0.233
22	bstw08_c	0.320	-0.192	0.335	-0.242	-0.577
23	bstw17_c	0.248	-0.284	-0.516	-0.186	0.330
24	bstw21_c	0.010	1.702	0.707	1.012	0.305
25	bevo07_c	0.036	0.540	0.002	0.253	0.251
26	bfkt02_c	-0.136	-0.122	0.412	-0.091	-0.503
27	bfkt04_c	-0.014	-0.298	0.102	0.417	0.315
28	bfkt07_c	0.628	0.168	0.638	0.286	-0.352
29	bgen04_c	0.268	-0.476	-0.343	-0.518	-0.175
30	bgen18_c	0.516	0.236	-0.439	-0.377	0.062

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
31	binf08_c	0.318	-0.010	-0.096	-0.111	-0.015
32	binf17_c	0.376	0.250	0.033	0.413	0.380
33	boek03_c	0.136	0.088	0.635	0.781	0.146
34	boek09_c	0.104	0.282	0.414	0.702	0.288
35	bstw04_c	0.060	0.436	0.057	0.693	0.636
36	bstw18_c	-0.240	-0.138	0.024	-0.081	-0.105
37	bevo01_c	0.012	0.224	0.535	0.320	-0.215
38	bfkt09_c	0.178	-0.326	-0.333	-0.505	-0.172
39	bfkt17_c	0.036	-0.182	0.265	-0.193	-0.458
40	bfkt19_c	-0.416	0.328	-0.128	0.140	0.268
41	bgen05_c	0.376	-0.162	0.013	-0.148	-0.161
42	bgen11_c	-0.672	-0.080	-0.178	0.196	0.374
43	binf07_c	0.156	0.220	-0.106	-0.002	0.104
44	binf19_c	-0.340	0.024	0.223	0.323	0.100
45	boek16_c	0.180	0.108	-0.121	-0.102	0.019
46	boek20_c	-0.700	-0.148	-0.174	-0.264	-0.090

	Item	Gender	Immigration background	Books		
		male vs female	without vs with	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
47	bstw13_c	-0.322	-0.372	-0.096	-0.111	-0.015
48	bstw16_c	-0.288	0.122	-0.351	-0.228	0.123
49	bevo02_c	0.054	0.032	0.466	0.020	-0.446
50	bfkt03_c	-0.526	0.176	-0.026	-0.159	-0.133
51	bfkt21_c	-0.126	0.190	0.531	0.492	-0.039
52	bgen21_c	-0.128	0.008	-0.001	-0.047	-0.046
53	binf05_c	-0.112	0.104	-0.453	-0.573	-0.120
54	binf20_c	0.076	0.354	-0.048	-0.078	-0.030
55	boek02_c	-0.028	-0.258	0.285	0.609	0.324
56	boek12_c	-0.416	-0.432	-0.013	-0.074	-0.061
57	boek13_c	0.026	-0.336	-0.126	0.645	0.771
58	bstw06_c	-0.052	-0.024	0.344	0.229	-0.115
59	bstw09_c	0.460	0.078	0.563	0.184	-0.379
60	bstw11_c	-0.318	-0.106	-0.021	-0.019	0.002
	main effect	-0.112	0.156	0.272	0.391	0.119