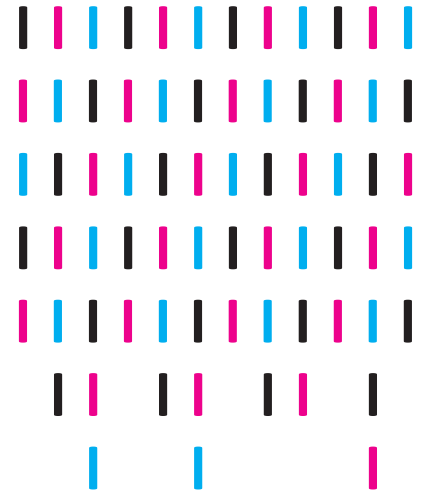




NEPS SURVEY PAPERS

Sven Rieger, Nicolas Hübner, and Wolfgang Wagner

NEPS TECHNICAL REPORT FOR BIO- LOGICAL COMPETENCE: SCALING RESULTS FOR THE ADDITIONAL STUDY THURINGIA

NEPS Survey Paper No. 41
Bamberg, June 2018

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 Thuringia

Sven Rieger, Nicolas Hübner, & Wolfgang Wagner

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

E-mail address of lead author:

sven.rieger@uni-tuebingen.de

Bibliographische Angaben:

Rieger, S., Hübner, N., & Wagner, W. (2018). *NEPS Technical Report for Biological Competence: Scaling Results for the Additional Study Thuringia* (NEPS Survey Paper No. 41). Bamberg: Leib-niz-Institute for Educational Trajectories, National Educational Panel Study. doi:10.5157/NEPS:SP41:1.0

We would like to thank Steffi Pohl and Claus. H. Carstensen for developing and providing standards for the technical reports. In addition, we would like to thank Eve Sarah Müller for her excellent support and valuable feedback.

NEPS Technical Report for Biological Competence: Scaling Results for the Additional Study Thuringia

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 Thuringia. 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 Baden-Wuerttemberg. In sum, 2,249 students took the biology test in these two waves. The biology test consisted of 126 items (distributed among seven booklets), 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 a subset of the items exhibited good item fit and measurement invariance across various groups. 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.

Keywords

item response theory, scaling, biological competence, scientific use file

Contents

Abstract	2
1 Introduction	4
2 Testing Biological Competence	4
3 Data	7
4 Analyses.....	7
4.1 Missing Responses.....	7
4.2 Scaling Model	7
4.3 Checking the Quality of the Scale.....	8
5 Results	9
5.1 Missing Responses.....	9
5.1.1 Missing responses per person.....	9
5.1.2 Missing responses per item	11
5.2 Parameter Estimates	11
5.2.1 Item parameters	11
5.2.2 Person parameters.....	29
5.2.3 Test targeting and reliability	29
5.3 Quality of the Test	31
5.3.1 Item fit.....	31
5.3.2 Differential item functioning.....	31
5.3.3 Rasch homogeneity.....	40
6 Discussion.....	40
7 Data in the Scientific Use File.....	41
References.....	42
Appendix.....	44

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.

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 two waves of the additional study Thuringia. Items in this test were developed in the Swiss EVAMAR II project (Evaluation der Maturitätsreform; Eberle et al., 2008) and used across two consecutive school years (2009/10 through 2010/11) 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 as well as further information about the test can be found on the NEPS website¹.

The present report draws strongly on previous technical reports such as Hübner, Rieger, & Wagner (2016), 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 II project 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 (SCR²; see Table 2). For MC items, the

¹ <https://www.neps-data.de/en-us/datacenter/dataanddocumentation/additionalstudythuringia.aspx>

² These items are sometimes scored as partial correct (e.g., 0 = false, 0.5 = partial correct, 1 = correct). This resulted in a ordinal response scale.

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 as well as booklets (for the content area of each item see Table S2 in the Appendix D).

Table 1

Content Areas of the Items on the Biology Test

Content area	Frequency
Cytology/ anatomy/ metabolism	42
Information processing/ characteristics/ immunobiology	21
Genetics/ developmental biology	21
Ecology	21
Systematics/ evolution	21
Total number of items	126

Number of Items by Content Area and Booklet	1	2	3	4	5	6	7
Cytology/ anatomy/ metabolism	6	6	6	6	6	6	6
Information processing/ characteristics/ immunobiology	3	3	3	3	3	3	3
Genetics/ developmental biology	3	3	3	3	3	3	3
Ecology	3	3	3	3	3	3	3
Systematics/ evolution	3	3	3	3	3	3	3
Total number of items	18	18	18	18	18	18	18

Table 2

Response Formats of the Items on the Biology Test

Response Format	Frequency
Single multiple choice	120
Short constructed response	6
Total number of items	126

Number of Items by Response Format and Booklet	1	2	3	4	5	6	7
Single multiple choice	18	17	17	17	17	17	17
Short constructed response	-	1	1	1	1	1	1
Total number of items	18	18	18	18	18	18	18

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.³ A total of 2,249 participants took the biology competence test: 1,367 in 2009/10 (Wave 1) and 882 in 2010/11 (Wave 2)⁴. All subjects gave at least one valid answer so that for every subject, a competence score was estimated.

4 Analyses

This chapter 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 (e.g., due to the different booklets). 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 partial credit model (PCM; Masters, 1982) was used and estimated in ConQuest 4.2 (Wu, Adams, Wilson, & Haldane, 2007).

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 (2012), whereas the data available in the SUF are described in Chapter 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.

3 <https://www.neps-data.de/en-us/datacenter/dataanddocumentation/additionalstudythuringia/documentation.aspx>

4 The dataset contains 2,260 persons.

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.⁵ We addressed the issue of measurement invariance by investigating test fairness for the variables gender, books at home (as a proxy for socioeconomic status; see Pohl and Carstensen, 2012 for a description of these variables), and wave (i.e., to which of the two waves do subjects belong?). 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, 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 (Muraki, 1990) with the TAM package in R (Robitzsch, Kiefer, & Wu, 2017; R Core Team, 2017).

⁵ 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 chapter, the key scaling results of the two waves of the additional study in Thuringia will be presented.

5.1 Missing Responses

In this subchapter, 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, 85.18 % of the participants did not produce any invalid responses. The maximum number of invalid responses was 34. Overall, 1.96% of the participants produced five or more invalid responses.

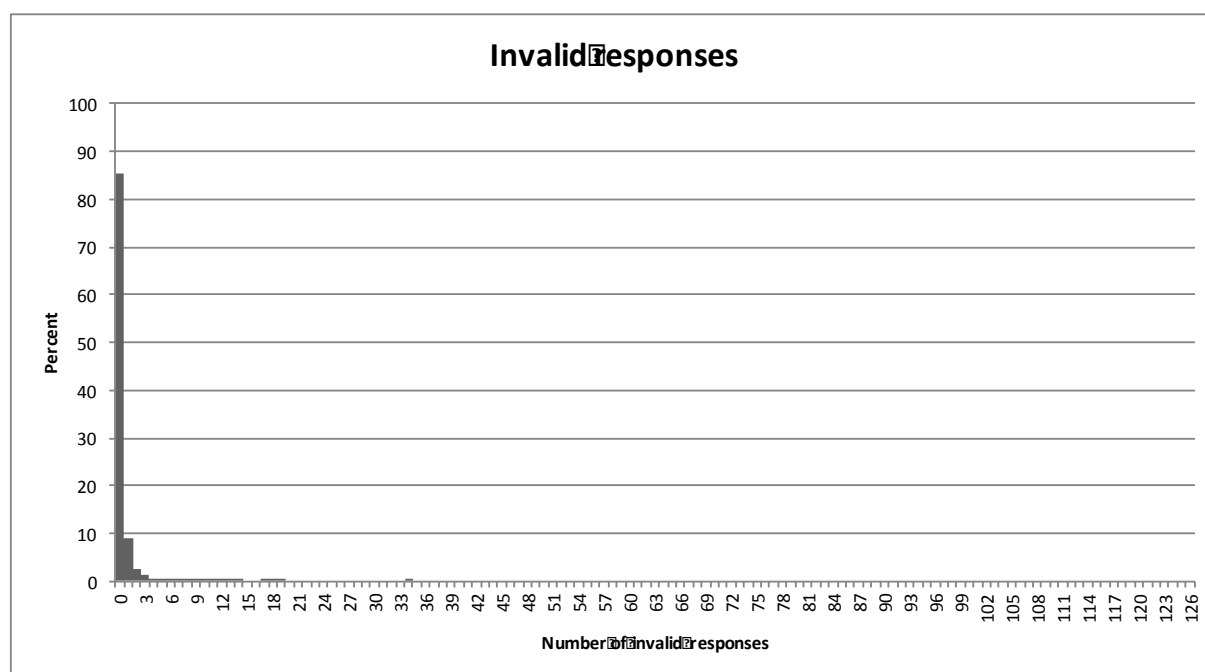


Figure 1. Number of invalid responses per person.

The largest source of missing responses on this test was the *omission of items*. As can be seen in Figure 2, 24.63% of the participants skipped at least one item. 4.19% of the participants omitted five or more items.

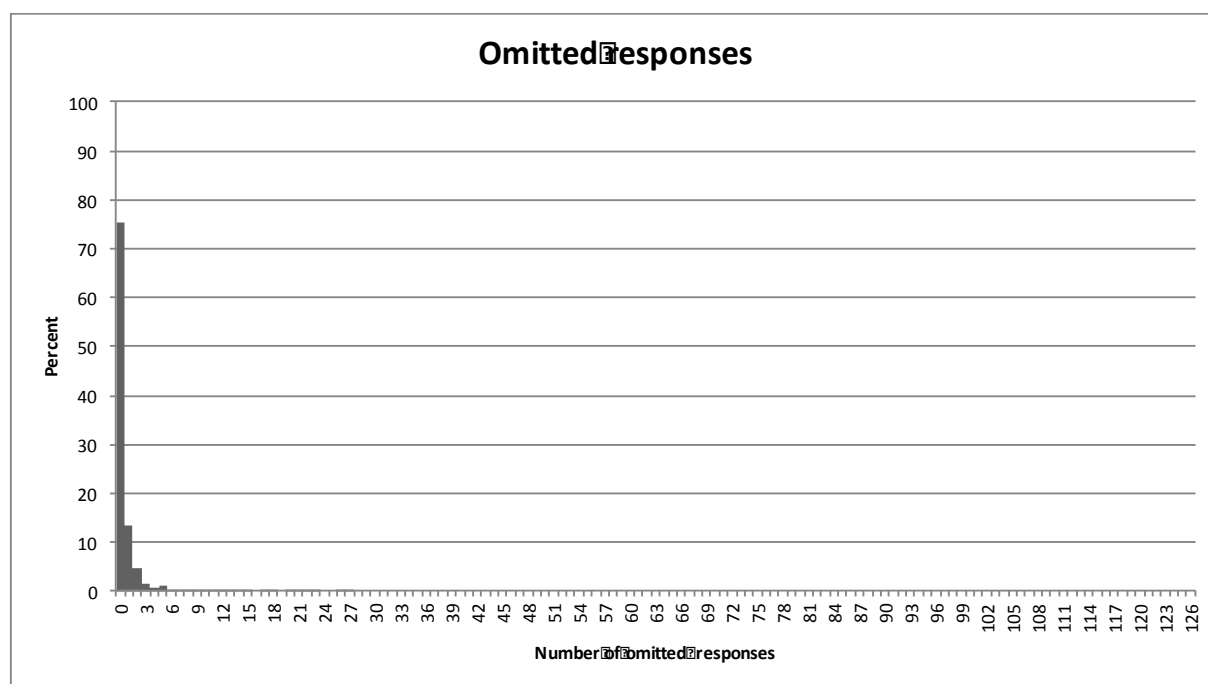


Figure 2. Number of omitted responses per person.

By definition, every item after the last item that was completed is labeled *not reached*. As Figure 3 shows, most participants (93.5%) reached the end of the test. Only 0.29% did not reach the fifth last item.

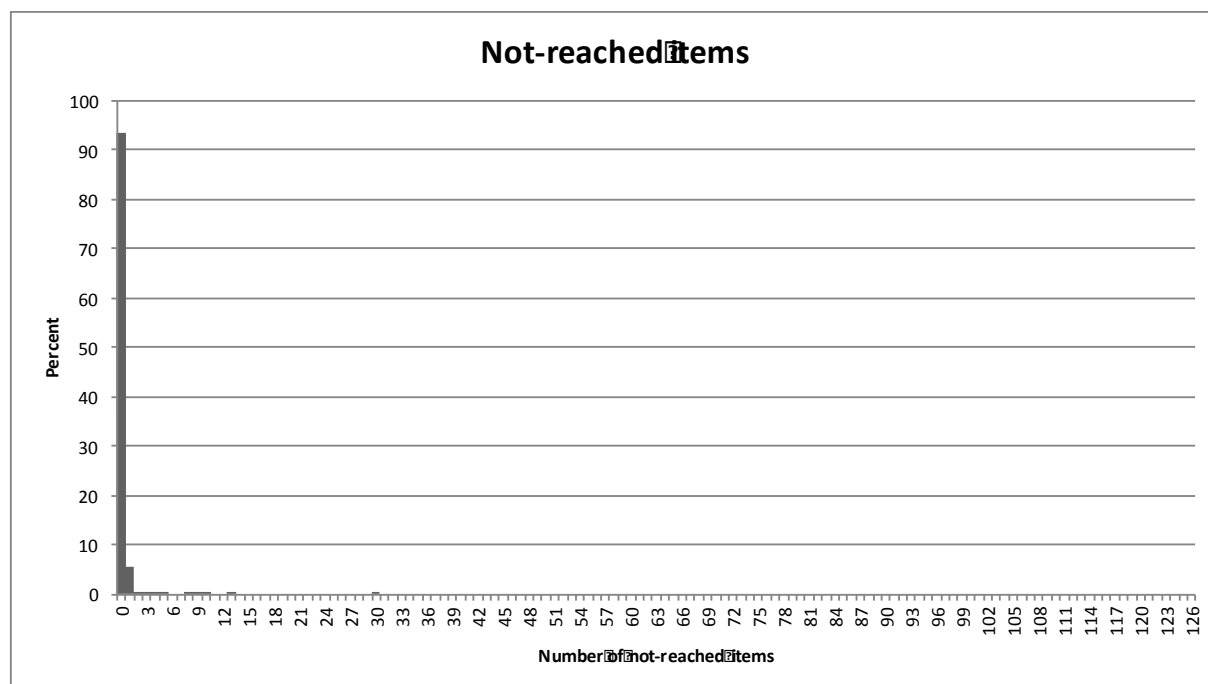


Figure 3. Number of not-reached items per person.

Overall, 93.5% of the participants had no items that were missing by *non-participation*. Only 0.5% (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 1.19 (SD = 3.01) missing responses. Moreover, 60.49% of the participants had no missing responses at all. Only 6.73% of the participants had five or more missing responses.

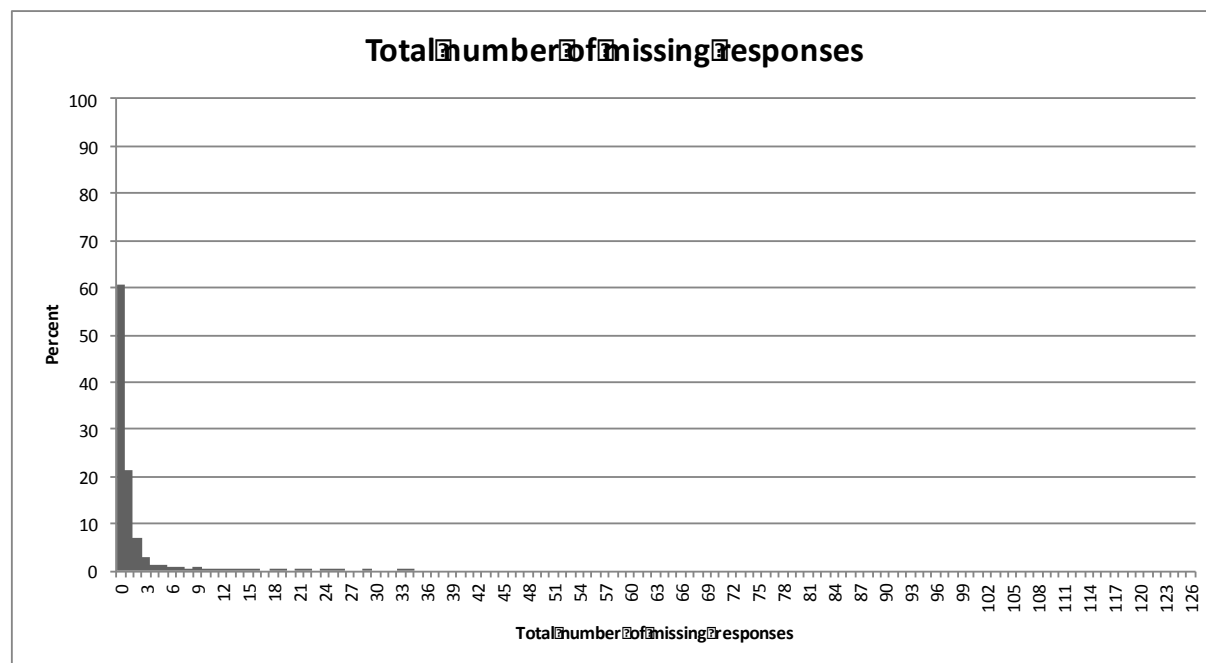


Figure 4. Total number of missing responses.

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 4.9% of the participants failed to reach items (column 7). However, with about 93.5% of the participants reaching the end of the test, the test can hardly be described as too long. Overall, 5 out of the 126 items had omission rates exceeding 2% (column 6). Items with open response format (short constructed responses) *bstw21s_c* (omitted by 2.5% of the participants), *boek21_c* (2.2%), *binf21s_c* (4.6%), *bfkt21_c* (2.3%) and *bevo21s_c* (2.6%) were the most noticeable. Overall, the percentage of invalid responses per item (column 5) was low (the maximum was 2.2% for item *bevo11_c*). The percentage of items that were missing by non-participation (column 8) was very low (the maximum was 0.5%). The percentage of missing by designs per items is displayed in column 8. The percentages ranged from 70.6% to 76.9%.

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 7.1% to 97.1% with an average of 48.29% (SD = 19.53%) 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 -3.603 (item bstw02_c) to 3.418 (item bstw21s_c) logits with an average difficulty of 0.058 logits ($SD = 1.024$). Altogether, the item difficulties were adequate. The step parameters for the polytomous items are displayed in Table 5. The GPCM discrimination parameters ranged from 0.019 to 3.348 (see again Table 4). The items bevo15_c, bevo03_c, bfkt11_c, bevo12_c, bfkt15_c, bgen02_c, and bgen09_c had a negative discrimination, paradoxically indicating that students with lower ability had a higher probability of solving the item. Therefore, after we rechecked the coding procedure, those items were excluded from further analyses (see Table S1 in Appendix C).

Table 3

Missing Values

	Item	Booklet	Position in the test	Number of valid responses	Percentage of invalid responses	Percentage of omitted responses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
1	bevo01_c	1,7	1 / 19	625	0.3	0.1	0.0	0.5	71.4
2	bevo08_c	1,7	2 / 20	587	0.1	2.0	0.0	0.5	71.4
3	bevo15_c	1,7	3 / 21	612	0.4	0.7	0.0	0.5	71.4
4	bfkt02_c	1,7	4 / 22	630	0.2	0.0	0.0	0.5	71.4
5	bfkt09_c	1,7	5 / 23	622	0.0	0.5	0.0	0.5	71.4
6	bfkt16_c	1,7	6 / 24	616	0.1	0.8	0.0	0.5	71.4
7	bgen03_c	1,7	7 / 25	614	0.3	0.6	0.0	0.5	71.4
8	bgen10_c	1,7	8 / 26	617	0.3	0.5	0.0	0.5	71.4
9	bgen17_c	1,7	9 / 27	620	0.2	0.5	0.0	0.5	71.4
10	binf04_c	1,7	10 / 28	629	0.0	0.2	0.0	0.5	71.4
11	binf11_c	1,7	11 / 29	622	0.3	0.3	0.0	0.5	71.4
12	binf18_c	1,7	12 / 30	615	0.4	0.5	0.0	0.5	71.4
13	boek05_c	1,7	13 / 31	619	0.1	0.6	0.0	0.5	71.4
14	boek12_c	1,7	14 / 32	626	0.2	0.2	0.0	0.5	71.4

	Item	Booklet	Position in the test	Number of valid responses	Percentage of invalid responses	Percentage of omitted responses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
15	boek19_c	1,7	15 / 33	627	0.2	0.1	0.1	0.5	71.4
16	bstw06_c	1,7	16 / 34	613	0.3	0.6	0.2	0.5	71.4
17	bstw13_c	1,7	17 / 35	625	0.1	0.1	0.2	0.5	71.4
18	bstw20_c	1,7	18 / 36	624	0.1	0.2	0.2	0.5	71.4
19	bevo02_c	1,2	19 / 1	622	0.3	0.3	-	0.5	71.4
20	bevo09_c	1,2	20 / 2	608	0.8	0.4	-	0.5	71.4
21	bevo16_c	1,2	21 / 3	622	0.4	0.2	-	0.5	71.4
22	bfkt03_c	1,2	22 / 4	626	0.4	0.0	-	0.5	71.4
23	bfkt10_c	1,2	23 / 5	629	0.1	0.1	-	0.5	71.4
24	bfkt17_c	1,2	24 / 6	610	0.3	0.8	-	0.5	71.4
25	bgen04_c	1,2	25 / 7	625	0.1	0.4	-	0.5	71.4
26	bgen11_c	1,2	26 / 8	618	0.3	0.4	-	0.5	71.4
27	bgen18_c	1,2	27 / 9	618	0.1	0.6	-	0.5	71.4
28	binf05_c	1,2	28 / 10	615	0.8	0.0	-	0.5	71.4
29	binf12_c	1,2	29 / 11	617	0.5	0.3	-	0.5	71.4
30	binf19_c	1,2	30 / 12	608	0.4	0.8	-	0.5	71.4

	Item	Booklet	Position in the test	Number of valid responses	Percentage of invalid responses	Percentage of omitted responses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
31	boek06_c	1,2	31 / 13	594	1.5	0.3	-	0.5	71.4
32	boek13_c	1,2	32 / 14	612	0.0	1.0	-	0.5	71.4
33	boek20_c	1,2	33 / 15	615	0.6	0.3	-	0.5	71.4
34	bstw07_c	1,2	34 / 16	615	0.3	0.6	-	0.5	71.4
35	bstw14_c	1,2	35 / 17	617	0.8	-	-	0.5	71.4
36	bstw21s_c	1,2	36 / 18	370	1.2	2.5	4.9	0.5	74.6
37	bevo03_c	2,3	19 / 1	629	0.0	0.4	-	0.5	71.3
38	bevo10_c	2,3	20 / 2	622	-	0.7	-	0.5	71.3
39	bevo17_c	2,3	21 / 3	623	0.4	0.3	-	0.5	71.3
40	bfkt04_c	2,3	22 / 4	627	0.4	0.1	-	0.5	71.3
41	bfkt11_c	2,3	23 / 5	620	0.1	0.7	-	0.5	71.3
42	bfkt18_c	2,3	24 / 6	631	0.1	0.2	-	0.5	71.3
43	bgen05_c	2,3	25 / 7	624	0.1	0.5	-	0.5	71.3
44	bgen12_c	2,3	26 / 8	629	0.2	0.2	-	0.5	71.3
45	bgen19_c	2,3	27 / 9	603	0.0	1.5	-	0.5	71.3
46	binf06_c	2,3	28 / 10	625	0.5	0.0	-	0.5	71.3

	Item	Booklet	Position in the test	Number of valid responses	Percentage of invalid responses	Percentage of omitted responses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
47	binf13_c	2,3	29 / 11	627	0.0	0.4	0.0	0.5	71.3
48	binf20_c	2,3	30 / 12	603	0.9	0.6	0.0	0.5	71.3
49	boek07_c	2,3	31 / 13	618	0.2	0.7	0.0	0.5	71.3
50	boek14_c	2,3	32 / 14	604	0.4	1.1	0.0	0.5	71.3
51	boek21_c	2,3	33 / 15	508	0.4	2.2	0.0	0.5	74.4
52	bstw01_c	2,3	34 / 16	627	0.0	0.3	0.1	0.5	71.3
53	bstw08_c	2,3	35 / 17	624	0.4	0.0	0.1	0.5	71.3
54	bstw15_c	2,3	36 / 18	627	-	0.1	0.4	0.5	71.3
55	bevo04_c	3,4	19 / 1	630	0.1	1.0	-	0.5	70.6
56	bevo11_c	3,4	20 / 2	596	2.2	0.4	-	0.5	70.6
57	bevo18_c	3,4	21 / 3	642	0.5	0.0	-	0.5	70.6
58	bfkt05_c	3,4	22 / 4	640	0.1	0.5	-	0.5	70.6
59	bfkt12_c	3,4	23 / 5	647	-	0.3	-	0.5	70.6
60	bfkt19_c	3,4	24 / 6	647	0.2	0.1	-	0.5	70.6
61	bgen06_c	3,4	25 / 7	636	0.7	0.1	-	0.5	70.6
62	bgen13_c	3,4	26 / 8	643	0.3	0.2	-	0.5	70.6

	Item	Booklet	Position in the test	Number of valid responses	Percentage of invalid responses	Percentage of omitted responses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
63	bgen20_c	3,4	27 / 9	639	0.2	0.4	0.0	0.5	70.6
64	binf07_c	3,4	28 / 10	646	0.2	0.1	0.0	0.5	70.6
65	binf14_c	3,4	29 / 11	645	0.2	0.1	0.0	0.5	70.6
66	binf21s_c	3,4	30 / 12	378	1.2	4.6	0.0	0.5	76.9
67	boek01_c	3,4	31 / 13	638	0.2	0.5	0.0	0.5	70.6
68	boek08_c	3,4	32 / 14	632	0.2	0.8	0.0	0.5	70.6
69	boek15_c	3,4	33 / 15	644	0.1	0.3	0.1	0.5	70.6
70	bstw02_c	3,4	34 / 16	651	-	0.0	0.1	0.5	70.6
71	bstw09_c	3,4	35 / 17	642	0.4	-	0.1	0.5	70.6
72	bstw16_c	3,4	36 / 18	647	-	-	0.3	0.5	70.6
73	bevo05_c	4,5	19 / 1	643	0.2	0.3	-	0.5	70.6
74	bevo12_c	4,5	20 / 2	636	0.1	0.7	-	0.5	70.6
75	bevo19_c	4,5	21 / 3	630	0.1	0.9	-	0.5	70.6
76	bfkt06_c	4,5	22 / 4	644	0.1	0.3	-	0.5	70.6
77	bfkt13_c	4,5	23 / 5	635	0.0	0.8	-	0.5	70.6
78	bfkt20_c	4,5	24 / 6	653	-	0.0	-	0.5	70.6
79	bgen07_c	4,5	25 / 7	643	0.1	0.4	-	0.5	70.6

	Item	Booklet	Position in the test	Number of valid re-sponses	Percentage of invalid re-sponses	Percentage of omitted re-sponses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
80	bgen14_c	4,5	26 / 8	641	0.2	0.4	-	0.5	70.6
81	bgen21s_c	4,5	27 / 9	567	0.4	1.5	-	0.5	72.6
82	binf01_c	4,5	28 / 10	629	-	1.1	-	0.5	70.6
83	binf08_c	4,5	29 / 11	643	-	0.5	-	0.5	70.6
84	binf15_c	4,5	30 / 12	641	0.2	0.4	-	0.5	70.6
85	boek02_c	4,5	31 / 13	641	0.5	0.1	-	0.5	70.6
86	boek09_c	4,5	32 / 14	640	0.1	0.5	0.0	0.5	70.6
87	boek16_c	4,5	33 / 15	642	0.2	0.3	0.0	0.5	70.6
88	bstw03_c	4,5	34 / 16	642	0.1	0.3	0.1	0.5	70.6
89	bstw10_c	4,5	35 / 17	627	-	1.1	0.1	0.5	70.6
90	bstw17_c	4,5	36 / 18	646	0.0	0.2	0.1	0.5	70.6
91	bevo06_c	5,6	19 / 1	596	0.3	2.0	-	0.5	70.8
92	bevo13_c	5,6	20 / 2	618	0.2	1.2	-	0.5	70.8
93	bevo20_c	5,6	21 / 3	613	0.8	0.8	-	0.5	70.8
94	bfkt07_c	5,6	22 / 4	634	0.0	0.6	-	0.5	70.8
95	bfkt14_c	5,6	23 / 5	609	1.5	0.2	-	0.5	70.8
96	bfkt21_c	5,6	24 / 6	512	0.5	2.3	-	0.5	74.1

	Item	Booklet	Position in the test	Number of valid responses	Percentage of invalid responses	Percentage of omitted responses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
97	bgen01_c	5,6	25 / 7	626	0.7	0.3	-	0.5	70.8
98	bgen08_c	5,6	26 / 8	621	0.8	0.4	-	0.5	70.8
99	bgen15_c	5,6	27 / 9	638	0.0	0.4	-	0.5	70.8
100	binf02_c	5,6	28 / 10	635	0.0	0.6	-	0.5	70.8
101	binf09_c	5,6	29 / 11	621	0.8	0.4	0.0	0.5	70.8
102	binf16_c	5,6	30 / 12	625	0.3	0.7	0.0	0.5	70.8
103	boek03_c	5,6	31 / 13	645	0.0	0.1	0.0	0.5	70.8
104	boek10_c	5,6	32 / 14	638	0.2	0.2	0.0	0.5	70.8
105	boek17_c	5,6	33 / 15	633	0.4	0.2	0.0	0.5	70.8
106	bstw04_c	5,6	34 / 16	642	0.1	0.1	0.0	0.5	70.8
107	bstw11_c	5,6	35 / 17	627	0.5	0.4	0.1	0.5	70.8
108	bstw18_c	5,6	36 / 18	640	0.1	0.0	0.2	0.5	70.8
109	bevo07_c	6,7	19 / 1	595	0.8	0.9	-	0.5	71.5
110	bevo14_c	6,7	20 / 2	619	-	0.6	-	0.5	71.5
111	bevo21s_c	6,7	21 / 3	464	0.6	2.6	-	0.5	75.8
112	bfkt01_c	6,7	22 / 4	621	-	0.5	-	0.5	71.5
113	bfkt08_c	6,7	23 / 5	592	-	1.8	-	0.5	71.5

Item	Booklet	Position in the test	Number of valid re-sponses	Percentage of invalid re-sponses	Percentage of omitted re-sponses	Percentage of not-reached responses	Percentage of missing by non-participation	Percentage of missing by design
114 bfkt15_c	6,7	24 / 6	611	-	0.9	0.0	0.5	71.5
115 bgen02_c	6,7	25 / 7	615	0.1	0.6	0.1	0.5	71.5
116 bgen09_c	6,7	26 / 8	594	-	1.6	0.1	0.5	71.5
117 bgen16_c	6,7	27 / 9	606	0.5	0.5	0.1	0.5	71.5
118 binf03_c	6,7	28 / 10	611	0.0	0.8	0.1	0.5	71.5
119 binf10_c	6,7	29 / 11	618	0.1	0.4	0.1	0.5	71.5
120 binf17_c	6,7	30 / 12	612	0.4	0.4	0.1	0.5	71.5
121 boek04_c	6,7	31 / 13	610	-	0.8	0.1	0.5	71.5
122 boek11_c	6,7	32 / 14	618	-	0.5	0.1	0.5	71.5
123 boek18_c	6,7	33 / 15	622	0.1	0.2	0.1	0.5	71.5
124 bstw05_c	6,7	34 / 16	604	0.6	0.5	0.2	0.5	71.5
125 bstw12_c	6,7	35 / 17	620	0.1	0.2	0.2	0.5	71.5
126 bstw19_c	6,7	36 / 18	618	0.1	0.1	0.4	0.5	71.5

Table 4

Item Parameters of the Biology Test

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination-2 PL
1	bevo01_c	75.52	-1.193	0.097	0.98	-0.4	0.31	1.166
2	bevo08_c	28.79	0.938	0.095	1.01	0.2	0.23	0.606
3	bevo15_c	-	-	-	-	-	-	-
4	bfkt02_c	51.75	-0.078	0.084	0.98	-1.1	0.34	1.308
5	bfkt09_c	45.18	0.201	0.085	1.01	0.5	0.28	0.750
6	bfkt16_c	62.99	-0.566	0.088	0.97	-1.2	0.37	1.533
7	bgen03_c	54.07	-0.182	0.085	1.05	2.5	0.19	0.255
8	bgen10_c	62.07	-0.521	0.087	0.99	-0.4	0.32	0.984
9	bgen17_c	53.87	-0.170	0.085	1.06	3.0	0.16	0.183
10	binf04_c	31.80	0.805	0.090	1.02	0.6	0.22	0.667
11	binf11_c	66.56	-0.735	0.089	0.95	-1.5	0.41	1.925
12	binf18_c	41.79	0.345	0.086	1.04	1.8	0.22	0.428
13	boek05_c	59.94	-0.430	0.086	1.01	0.5	0.26	0.622
14	boek12_c	51.12	-0.051	0.084	0.97	-1.7	0.36	1.274
15	boek19_c	89.95	-2.294	0.136	1.00	0.0	0.22	1.344
16	bstw06_c	35.07	0.646	0.089	0.96	-1.4	0.39	1.629

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
17	bstw13_c	65.28	-0.673	0.088	0.99	-0.4	0.32	1.207
18	bstw20_c	45.51	0.184	0.085	1.03	1.7	0.22	0.555
19	bevo02_c	70.58	-0.921	0.092	1.01	0.2	0.29	0.914
20	bevo09_c	56.91	-0.299	0.086	1.00	0.1	0.30	0.889
21	bevo16_c	34.89	0.658	0.088	0.97	-1.0	0.36	1.309
22	bfkt03_c	80.51	-1.485	0.104	0.96	-0.7	0.34	2.189
23	bfkt10_c	61.05	-0.474	0.086	0.98	-0.7	0.31	1.280
24	bfkt17_c	33.11	0.747	0.090	1.00	-0.1	0.30	1.066
25	bgen04_c	34.24	0.691	0.088	0.99	-0.2	0.29	0.977
26	bgen11_c	24.92	1.169	0.097	0.98	-0.3	0.29	1.077
27	bgen18_c	17.64	1.625	0.109	0.99	-0.2	0.27	1.208
28	binf05_c	53.33	-0.139	0.085	1.00	0.2	0.29	0.815
29	binf12_c	45.87	0.177	0.085	1.03	1.7	0.21	0.439
30	binf19_c	38.65	0.488	0.087	0.96	-1.7	0.40	1.503
31	boek06_c	31.31	0.829	0.093	1.03	0.7	0.21	0.484
32	boek13_c	37.75	0.534	0.088	0.98	-0.7	0.32	1.148
33	boek20_c	54.96	-0.210	0.085	1.00	0.2	0.29	0.959
34	bstw07_c	42.60	0.318	0.086	0.99	-0.4	0.32	1.061

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
35	bstw14_c	70.34	-0.914	0.092	1.01	0.3	0.23	0.673
36	bstw21s_c	n.a.	3.418	0.309	0.96	-0.3	0.32	1.404
37	bevo03_c	-	-	-	-	-	-	-
38	bevo10_c	36.98	0.556	0.087	1.02	0.8	0.22	0.399
39	bevo17_c	43.98	0.243	0.085	1.04	2.2	0.20	0.277
40	bfkt04_c	53.75	-0.163	0.084	0.97	-1.5	0.36	1.254
41	bfkt11_c	-	-	-	-	-	-	-
42	bfkt18_c	61.49	-0.501	0.086	1.00	-0.2	0.30	1.001
43	bgen05_c	27.56	1.015	0.093	0.95	-1.2	0.40	1.817
44	bgen12_c	59.30	-0.401	0.085	1.03	1.4	0.21	0.418
45	bgen19_c	28.19	0.975	0.094	1.01	0.2	0.23	0.558
46	binf06_c	59.52	-0.415	0.085	1.03	1.2	0.24	0.549
47	binf13_c	37.64	0.527	0.086	1.01	0.3	0.26	0.650
48	binf20_c	41.79	0.346	0.087	0.99	-0.6	0.29	0.961
49	boek07_c	29.77	0.900	0.092	1.03	0.7	0.18	0.299
50	boek14_c	37.42	0.541	0.088	0.98	-0.9	0.35	1.170
51	boek21_c	73.62	-1.054	0.105	1.04	0.9	0.13	0.200
52	bstw01_c	29.98	0.887	0.091	1.04	0.9	0.18	0.343

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
53	bstw08_c	67.15	-0.765	0.089	0.97	-1.0	0.36	1.473
54	bstw15_c	38.12	0.504	0.086	1.02	0.7	0.24	0.538
55	bevo04_c	13.97	1.900	0.118	1.02	0.2	0.15	0.343
56	bevo11_c	17.95	1.578	0.110	1.02	0.4	0.15	0.519
57	bevo18_c	55.14	-0.229	0.083	1.00	0.1	0.29	0.870
58	bfkt05_c	28.44	0.963	0.091	1.01	0.1	0.26	0.823
59	bfkt12_c	73.57	-1.086	0.093	0.99	-0.2	0.30	1.190
60	bfkt19_c	70.63	-0.933	0.090	0.95	-1.3	0.40	2.057
61	bgen06_c	47.48	0.096	0.083	1.03	1.6	0.22	0.505
62	bgen13_c	28.62	0.957	0.091	1.02	0.6	0.21	0.536
63	bgen20_c	23.00	1.263	0.098	1.03	0.6	0.17	0.375
64	binf07_c	82.66	-1.645	0.107	0.97	-0.4	0.33	1.724
65	binf14_c	56.28	-0.275	0.083	1.02	1.0	0.25	0.494
66	binf21s_c	n.a.	1.890	0.175	0.95	-0.7	0.40	1.627
67	boek01_c	21.47	1.356	0.100	1.03	0.5	0.17	0.232
68	boek08_c	43.20	0.281	0.084	1.03	1.3	0.22	0.379
69	boek15_c	36.96	0.555	0.086	1.01	0.5	0.26	0.575
70	bstw02_c	97.08	-3.624	0.234	0.98	0.0	0.21	3.348

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
71	bstw09_c	88.32	-2.122	0.126	0.97	-0.3	0.33	2.610
72	bstw16_c	67.85	-0.795	0.088	0.95	-1.6	0.42	2.430
73	bevo05_c	69.05	-0.840	0.089	0.99	-0.3	0.31	0.826
74	bevo12_c	-	-	-	-	-	-	-
75	bevo19_c	7.14	2.678	0.157	1.01	0.1	0.12	0.418
76	bfkt06_c	43.94	0.266	0.084	0.98	-0.9	0.35	1.078
77	bfkt13_c	29.13	0.945	0.091	1.00	-0.1	0.27	0.909
78	bfkt20_c	77.18	-1.279	0.097	1.03	0.6	0.17	0.472
79	bgen07_c	44.63	0.234	0.083	1.04	1.9	0.22	0.393
80	bgen14_c	70.83	-0.931	0.091	0.97	-0.8	0.35	1.334
81	bgen21s_c	n.a.	0.987	0.163	0.98	-0.3	0.27	0.784
82	binf01_c	34.98	0.661	0.088	1.01	0.4	0.27	0.602
83	binf08_c	41.37	0.376	0.084	0.97	-1.2	0.35	1.178
84	binf15_c	58.35	-0.352	0.084	1.02	0.9	0.25	0.579
85	boek02_c	77.54	-1.298	0.098	0.96	-0.8	0.36	1.886
86	boek09_c	52.34	-0.093	0.083	0.97	-1.9	0.38	1.479
87	boek16_c	42.06	0.346	0.084	1.00	0.0	0.31	0.959
88	bstw03_c	37.54	0.546	0.086	1.06	2.4	0.12	0.031

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
89	bstw10_c	24.72	1.176	0.096	1.04	0.9	0.14	0.135
90	bstw17_c	63.47	-0.577	0.086	0.94	-2.1	0.42	1.768
91	bevo06_c	30.37	0.872	0.093	1.05	1.3	0.13	0.019
92	bevo13_c	23.95	1.216	0.098	1.02	0.4	0.19	0.536
93	bevo20_c	37.85	0.526	0.087	1.00	0.2	0.28	0.937
94	bfkt07_c	31.23	0.835	0.090	1.00	0.0	0.30	1.023
95	bfkt14_c	80.79	-1.495	0.106	0.98	-0.4	0.30	1.530
96	bfkt21_c	50.20	0.026	0.093	0.99	-0.5	0.31	1.119
97	bgen01_c	49.20	0.030	0.084	1.03	1.7	0.21	0.339
98	bgen08_c	46.54	0.143	0.085	0.99	-0.6	0.33	1.203
99	bgen15_c	78.21	-1.342	0.099	1.03	0.6	0.15	0.198
100	binf02_c	56.85	-0.293	0.084	1.03	1.5	0.22	0.494
101	binf09_c	30.11	0.885	0.091	1.01	0.2	0.27	0.991
102	binf16_c	36.64	0.577	0.087	1.05	1.9	0.15	0.067
103	boek03_c	86.36	-1.930	0.118	0.98	-0.3	0.32	2.075
104	boek10_c	64.89	-0.647	0.087	0.97	-1.2	0.37	1.395
105	boek17_c	23.70	1.232	0.097	0.96	-0.8	0.37	1.896
106	bstw04_c	89.25	-2.208	0.130	0.96	-0.4	0.37	2.885

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
107	bstw11_c	33.17	0.736	0.089	0.96	-1.4	0.39	1.622
108	bstw18_c	59.22	-0.395	0.084	1.01	0.3	0.26	0.727
109	bevo07_c	34.62	0.681	0.090	0.98	-0.5	0.33	1.117
110	bevo14_c	58.32	-0.351	0.086	1.02	1.0	0.24	0.591
111	bevo21s_c	n.a.	-0.605	0.102	1.02	0.6	0.22	0.296
112	bfkt01_c	25.12	1.160	0.096	1.00	0.0	0.24	0.743
113	bfkt08_c	41.05	0.389	0.088	0.99	-0.3	0.32	1.116
114	bfkt15_c	-	-	-	-	-	-	-
115	bgen02_c	-	-	-	-	-	-	-
116	bgen09_c	-	-	-	-	-	-	-
117	bgen16_c	30.36	0.880	0.092	0.99	-0.3	0.29	1.070
118	binf03_c	51.39	-0.052	0.085	1.02	0.9	0.26	0.572
119	binf10_c	72.17	-1.001	0.094	0.98	-0.5	0.32	1.134
120	binf17_c	51.80	-0.072	0.085	0.98	-0.8	0.33	0.964
121	boek04_c	71.80	-0.980	0.094	1.00	0.1	0.28	0.998
122	boek11_c	54.85	-0.199	0.085	1.05	2.5	0.17	0.201
123	boek18_c	24.28	1.207	0.097	1.03	0.6	0.18	0.484
124	bstw05_c	52.81	-0.113	0.086	1.01	0.6	0.27	0.706

Item		Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
125	bstw12_c	76.61	-1.245	0.099	0.95	-1.0	0.42	2.270
126	bstw19_c	54.21	-0.173	0.085	0.97	-1.6	0.38	1.250

Note. Difficulty = Item difficulty / location parameter, SE = Standard error of item difficulty / location parameter, WMNSQ = Weighted mean square, t = t-value for WMNSQ. Items 3, 37, 41, 74, 114, 115, and 116 were excluded from the analyses due to an unsatisfactory item fit. Percent correct scores are not informative for polytomous CMC item scores. These are denoted by n.a.

Table 5

Step Parameters (and Standard Errors) of the Polytomous Item

Item	Step 1 (SE)	Step 2
bstw21s_c	-0.322 (0.192)	0.322
binf21s_c	-0.157 (0.129)	0.157
bgen21s_c	-1.509 (0.096)	1.509
bevo21s_c	3.415 (0.381)	-3.415

Note. The last step parameter is not estimated and has, thus, no standard error because it is a constrained parameter for model identification.

5.2.2 Person parameters

The person parameters were estimated as WLEs (Pohl & Carstensen, 2012). A description of the data in the SUF can be found in Chapter 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 all parts of the ability distribution very well. Hence, the test can measure person abilities in the low to high regions comparably well.

The mean of the ability distribution was constrained to be zero, and its variance was estimated to be 0.25⁶. The reliability of the test (EAP/PV reliability = .61, WLE reliability = .61) was acceptable, but not good. This might be related to the suboptimal discrimination of test items.

⁶ Seven items (i.e., bevo15_c, bevo03_c, bfkt11_c, bevo12_c, bfkt15_c, bgen02_c, and bgen09_c) were excluded due to negative item discriminations (see Table S1 in the Appendix).

Scale (in logits)	Person ability	Item difficulty
1		35 71 52 63
		26 53
	X	64
		60 88 101 116
	X	25 85 108
	X	
	XX	40
	XX	2 42 55 59 73 77
	X	46 49 87 97 110
	XX	9 30 90
0	XX	23 103
	XXXX	24 105
	XXXX	15 20 78
	XXXXX	31 36 47 66 84 98
	XXXXXXX	29 44 51 89
	XXXXXX	
	XXXXXXXX	11 45 79 83 109
	XXXXXXXX	33 65
	XXXXXXXX	4 37 72 75
	XXXXXXXXXX	17 28 94
-1	XXXXXXXXXX	58
	XXXXXXXXXX	92 93
	XXXXXXXXXX	13 111
	XXXXXXXXXX	3 82 113 117
	XXXXXXXXXX	6 8 27 38 115 119
	XXXXXXXXXX	32 54
	XXXXXXXXXX	19 62 96
	XXXXXXXXXX	80 104 106
	XXXXXXX	12 41 43
	XXXX	7 22 39
-2	XXXXX	5 86
	XXXXX	100 107
	XXXX	16
	XXX	10 50 69
	XX	70
	XX	18 34 57 76
	X	114
	X	48 112
	X	56
	X	1
		118 74 81 95
		21 91
		61
		99
		14 67 68 102

Figure 5. Test targeting. The distribution of person abilities in the sample is depicted on the left-hand side, with each 'X' representing 12.7 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 0.94 (items bstw17_c) to 1.06 (items bgen17_c and bstw03_c), cf. column 5 of Table 4. Point-biserial correlations between the item scores and the total scores ranged from 0.12 (items bevo19_c and bstw03_c) to 0.42 (items bstw16_c, bstw17_c, and bstw12_c). Discriminations estimated in the generalized partial credit model with the TAM package in R ranged from 0.019 (item bevo06_c) to 3.348 (item bstw02_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, wave, and books (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. In sum, seven items exceeded this threshold for the current test. This should be considered when using the WLE estimates (see also Chapter 7) in conjunction with these grouping variables.

The table depicts the differences in the estimated item difficulties between the respective groups. “Male vs. Female”, for example, indicates the difference in difficulty $\beta_{\text{male}} - \beta_{\text{female}}$. 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, female participants did not differ from male participants (main effect = 0.020 logits, Cohen’s $d = 0.040$).⁷ Ten items (bfkt02_c, bgen04_c, bstw14_c, binf21s_c, bstw02_c, bgen21s_c, bgen15_c, bgen16_c, binf10_c, and boek18_c) showed a DIF greater than 0.6 logits. Two items (boek19_c and bstw21s_c) showed a very strong DIF exceeding 1 logit.

Wave: On average, participants in the two waves did not differ in their biological competence (1 vs 2: main effect = 0.004, Cohen’s $d = 0.008$). Two items (bstw21s_c and bevo19_c) showed a very strong DIF exceeding 1 logit.

Books: On average, participants with many books at home performed better on the biological competence test (0-200 vs 201-500: main effect = 0.135, Cohen’s $d = 0.270$; 0-200 vs > 500: main effect = 0.309, Cohen’s $d = 0.618$; 201-500 vs > 500: main effect = 0.174, Cohen’s $d = 0.348$). Eleven items (bevo01_c, bgen18_c, boek20_c, bstw07_c, boek14_c, bstw15_c, bevo04_c, bgen15_c, binf16_c, boek03_c, boek17_c) showed a DIF greater than 0.6 logits. Three items (bstw21s_c, binf21s_c, and bstw02_c) showed a very strong DIF exceeding 1 logit.

⁷ The variance of the Partial-Credit Model was used to estimate the effect size.

Table 6

Differential Item Functioning

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
1	bevo01_c	-0.244	0.024	0.043	0.605	0.562
2	bevo08_c	-0.182	0.040	-0.117	-0.310	-0.193
3	bevo15_c	-	-	-	-	-
4	bfkt02_c	-0.806	-0.022	0.038	0.347	0.309
5	bfkt09_c	-0.188	-0.056	-0.309	-0.562	-0.253
6	bfkt16_c	0.420	-0.124	0.201	0.368	0.167
7	bgen03_c	0.128	0.244	-0.317	-0.077	0.240
8	bgen10_c	0.322	0.254	0.084	-0.267	-0.351
9	bgen17_c	0.086	-0.138	-0.198	-0.505	-0.307
10	binf04_c	-0.242	0.012	0.023	-0.062	-0.085
11	binf11_c	-0.062	-0.130	-0.010	0.328	0.338
12	binf18_c	-0.298	-0.510	0.089	-0.191	-0.280
13	boek05_c	0.078	0.110	-0.023	0.102	0.125
14	boek12_c	-0.092	-0.268	-0.459	-0.002	0.457
15	boek19_c	-1.422	0.330	0.139	0.521	0.382
16	bstw06_c	-0.032	0.152	-0.186	-0.003	0.183

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
17	bstw13_c	0.182	0.288	-0.129	0.283	0.412
18	bstw20_c	-0.188	-0.096	-0.032	-0.507	-0.475
19	bevo02_c	-0.540	-0.390	-0.217	-0.020	0.197
20	bevo09_c	0.086	0.062	-0.328	0.138	0.466
21	bevo16_c	-0.326	-0.320	0.026	0.523	0.497
22	bfkt03_c	0.336	0.088	-0.041	0.167	0.208
23	bfkt10_c	0.020	0.286	0.222	0.369	0.147
24	bfkt17_c	0.128	-0.408	-0.314	-0.144	0.170
25	bgen04_c	0.670	0.178	0.045	0.008	-0.037
26	bgen11_c	-0.058	0.058	-0.115	0.319	0.434
27	bgen18_c	0.104	0.010	0.646	0.674	0.028
28	binf05_c	-0.462	-0.068	-0.452	-0.379	0.073
29	binf12_c	0.170	0.000	-0.159	0.000	0.159
30	binf19_c	0.206	0.030	-0.082	0.072	0.154
31	boek06_c	-0.094	-0.102	-0.098	-0.144	-0.046
32	boek13_c	-0.314	0.378	0.267	0.384	0.117
33	boek20_c	-0.512	0.038	-0.166	0.553	0.719
34	bstw07_c	-0.082	-0.222	0.218	0.617	0.399

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
35	bstw14_c	0.800	0.024	-0.299	-0.385	-0.086
36	bstw21s_c	-1.194	-1.034	-0.822	0.453	1.275
37	bevo03_c	-	-	-	-	-
38	bevo10_c	0.030	-0.044	-0.110	-0.031	0.079
39	bevo17_c	-0.234	0.114	0.004	-0.031	-0.035
40	bfkt04_c	-0.584	-0.192	0.263	0.221	-0.042
41	bfkt11_c	-	-	-	-	-
42	bfkt18_c	0.420	0.030	-0.234	-0.279	-0.045
43	bgen05_c	0.232	-0.100	0.393	0.426	0.033
44	bgen12_c	0.422	0.162	-0.276	-0.276	0.000
45	bgen19_c	0.352	-0.120	0.300	-0.024	-0.324
46	binf06_c	0.276	0.132	0.083	0.130	0.047
47	binf13_c	-0.028	0.020	-0.327	-0.019	0.308
48	binf20_c	0.210	0.116	0.365	0.010	-0.355
49	boek07_c	-0.046	0.134	0.260	0.058	-0.202
50	boek14_c	-0.384	0.326	-0.617	-0.097	0.520
51	boek21_c	-0.096	-0.204	0.211	0.050	-0.161
52	bstw01_c	0.120	-0.176	-0.357	-0.198	0.159

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
53	bstw08_c	-0.062	-0.176	0.011	0.367	0.356
54	bstw15_c	-0.070	-0.128	-0.474	-0.654	-0.180
55	bevo04_c	-0.050	0.208	-0.748	-0.482	0.266
56	bevo11_c	-0.318	0.208	-0.250	-0.536	-0.286
57	bevo18_c	0.432	0.106	0.258	0.184	-0.074
58	bfkt05_c	0.332	-0.158	0.444	-0.039	-0.483
59	bfkt12_c	0.252	0.408	0.341	-0.042	-0.383
60	bfkt19_c	-0.078	-0.004	0.266	0.499	0.233
61	bgen06_c	0.152	-0.174	0.258	0.313	0.055
62	bgen13_c	0.158	-0.168	-0.177	-0.387	-0.210
63	bgen20_c	-0.002	0.098	-0.064	-0.236	-0.172
64	binf07_c	0.470	-0.056	0.261	0.144	-0.117
65	binf14_c	-0.292	-0.194	0.068	-0.053	-0.121
66	binf21s_c	0.610	-0.184	1.643	0.871	-0.772
67	boek01_c	0.396	0.188	-0.326	-0.488	-0.162
68	boek08_c	-0.022	0.128	-0.266	-0.358	-0.092
69	boek15_c	0.098	0.242	0.028	0.210	0.182
70	bstw02_c	0.694	-0.326	1.596	-0.510	-2.106

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
71	bstw09_c	0.314	-0.254	0.376	-0.172	-0.548
72	bstw16_c	0.416	0.358	0.477	0.060	-0.417
73	bevo05_c	0.256	0.350	0.102	-0.006	-0.108
74	bevo12_c	-	-	-	-	-
75	bevo19_c	-0.436	-1.556	-0.474	-0.309	0.165
76	bfkt06_c	-0.146	-0.086	0.316	0.308	-0.008
77	bfkt13_c	0.100	0.040	0.453	-0.011	-0.464
78	bfkt20_c	0.350	-0.432	0.204	-0.085	-0.289
79	bgen07_c	0.210	0.332	-0.068	-0.580	-0.512
80	bgen14_c	0.156	0.310	0.124	-0.253	-0.377
81	bgen21s_c	0.672	0.018	0.257	0.475	0.218
82	binf01_c	-0.382	-0.112	-0.072	-0.066	0.006
83	binf08_c	0.326	0.174	0.065	0.103	0.038
84	binf15_c	0.264	-0.168	0.175	-0.091	-0.266
85	boek02_c	0.048	0.204	-0.144	0.027	0.171
86	boek09_c	0.052	0.228	0.090	0.231	0.141
87	boek16_c	-0.028	-0.198	-0.070	-0.018	0.052
88	bstw03_c	-0.082	-0.284	-0.175	-0.173	0.002

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
89	bstw10_c	0.264	-0.006	-0.032	-0.400	-0.368
90	bstw17_c	0.274	0.130	-0.154	-0.044	0.110
91	bevo06_c	0.072	-0.158	0.236	0.037	-0.199
92	bevo13_c	-0.430	0.188	-0.293	-0.088	0.205
93	bevo20_c	-0.516	0.088	0.266	0.548	0.282
94	bfkt07_c	0.512	0.118	-0.162	-0.147	0.015
95	bfkt14_c	-0.104	0.062	-0.378	-0.021	0.357
96	bfkt21_c	-0.032	-0.094	0.276	0.330	0.054
97	bgen01_c	-0.320	-0.224	-0.112	-0.011	0.101
98	bgen08_c	-0.472	0.486	-0.321	-0.054	0.267
99	bgen15_c	0.606	0.026	0.106	-0.763	-0.869
100	binf02_c	0.376	-0.150	-0.224	-0.322	-0.098
101	binf09_c	-0.234	0.050	-0.492	-0.415	0.077
102	binf16_c	0.006	0.116	-0.147	-0.757	-0.610
103	boek03_c	-0.030	0.036	0.471	0.642	0.171
104	boek10_c	-0.316	-0.056	0.064	0.014	-0.050
105	boek17_c	-0.556	-0.098	0.311	0.786	0.475
106	bstw04_c	0.020	-0.268	-0.057	0.300	0.357

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
107	bstw11_c	-0.392	0.296	-0.157	-0.095	0.062
108	bstw18_c	0.058	-0.048	-0.270	-0.240	0.030
109	bevo07_c	-0.196	0.160	-0.054	0.274	0.328
110	bevo14_c	0.230	-0.012	-0.017	-0.226	-0.209
111	bevo21s_c	-0.024	0.158	-0.242	-0.055	0.187
112	bfkt01_c	-0.124	0.084	0.272	0.202	-0.070
113	bfkt08_c	-0.368	-0.286	0.073	0.056	-0.017
114	bfkt15_c	-	-	-	-	-
115	bgen02_c	-	-	-	-	-
116	bgen09_c	-	-	-	-	-
117	bgen16_c	-0.694	-0.288	-0.245	-0.562	-0.317
118	binf03_c	0.324	0.058	-0.075	-0.327	-0.252
119	binf10_c	0.900	0.078	0.106	-0.184	-0.290
120	binf17_c	0.140	-0.144	0.105	0.087	-0.018
121	boek04_c	0.032	-0.142	0.034	-0.028	-0.062
122	boek11_c	-0.130	-0.192	-0.098	-0.427	-0.329
123	boek18_c	-0.956	0.336	0.528	0.510	-0.018
124	bstw05_c	0.018	-0.156	0.083	0.121	0.038

Item		Gender	Wave	Books		
		male vs female	1 vs 2	0-200 vs 201-500	0-200 vs > 500	201-500 vs > 500
125	bstw12_c	-0.094	0.004	-0.182	0.140	0.322
126	bstw19_c	-0.090	-0.014	-0.051	0.072	0.123
main effect		0.020	0.004	0.135	0.309	0.174

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 complex model including DIF was preferred only for the variable gender. 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 parsimonious models including only main effects were preferred over the ones containing the variables gender, wave and books.

Table 7

Comparison of Models With and Without DIF

DIF variable	Model	Number of parameters	AIC	BIC
Gender	main effect	125	89,527.30	89,696.30
	DIF	248	89,379.23	89,714.52
Wave	main effect	125	89,576.79	89,745.79
	DIF	248	89,664.67	89,999.96
Books	main effect	126	74,598.54	74,768.89
	DIF	372	74,746.22	75,249.16

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 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 0.019 (item bevo06_c) to 3.348 (item bstw02_c). Despite the empirical preference (according to the AIC, but not the BIC) for the generalized partial credit model (AIC = 89139.71, BIC = 90529.24, number of parameters = 243), the partial credit model (AIC = 89562.20, BIC = 90271.27, number of parameters = 124) 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 partial credit model was chosen as the scaling model.

6 Discussion

Descriptions and analyses presented in the previous chapters were aimed at documenting the quality of the biological competence test used in the additional study Thuringia. 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 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 126 items, of which 122 are scored as dichotomous variables with 0 (*incorrect response*) or 1 (*correct response*). Overall, 4 items were polytomous and in accordance with Pohl and 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, & Haldane, 2007). Appendix B provides an alternative syntax for use with the TAM package (Robitzsch, Kiefer, & Wu, 2017) in the software R (R Core Team, 2017).

Manifest biological competence scores are provided in the form of WLEs (b_sc1) along with their corresponding standard errors (b_sc2). As described in Chapter 5, these person estimates were derived from the joint scaling of all two 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; except items with negative discriminations in the 2PL were excluded (items bevo15_c, bevo03_c, bfkt11_c, bevo12_c, bfkt15_c, bgen02_c, and bgen09_c were excluded). Therefore, the delivered WLE is based on 119 items. 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 analyses that require one of the variables that showed DIF > 0.6 or 1 logits, we emphasize that (latent variable) models should be considered on the basis of 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.
- Eberle, F., Gehrler, 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.
- Hübner, N., Rieger, S., & Wagner, W. (2016). NEPS Technical Report for Biological Competence: Scaling Results for the Additional Study Baden-Württemberg (NEPS Survey Paper No.9). 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.
- 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 (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Robitzsch, A., Kiefer, T., & Wu, M. (2017). *TAM: Test analysis modules. R package version 2.7-56*. Retrieved from <https://CRAN.R-project.org/package=TAM>
- 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).
- 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., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalized item response modelling software*. Camberwell, AUS: ACER Press.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)*. Ottawa: National Defense Headquarters.

This paper used data from the National Educational Panel Study (NEPS): Additional Study Thuringia doi:10.5157/NEPS:TH:2.0.0. From 2008 to 2013, NEPS data were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

Appendix

Appendix A: ConQuest Syntax for generating WLE estimates in the Additional Study Thuringia

title Additional Study Thuringia, biological competence, Waves 1-2, Partial credit model;

datafile filename.dat;

format pid 1-7 responses 11-129;

labels << labels.nam;

codes 0,1,2;

score (0,1) (0,1) !item(1-34,36-62,64-76,78-106,108-119);

score (0,1,2) (0,.5,1) !item(35,63,77,107);

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 Thuringia

```
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 four polytomous items according to NEPS conventions
```

```
B[,3,][35] <- 1
```

```
B[,2,][35] <- 0.5
```

```
B[,3,][63] <- 1
```

```
B[,2,][63] <- 0.5
```

```
B[,3,][77] <- 1
```

```
B[,2,][77] <- 0.5
```

```
B[,3,][107] <- 1
```

```
B[,2,][107] <- 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
```

Appendix C: Item Parameters based on all Items

Table S1

Item Parameters of the English Test (all Items)

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
1	bevo01_c	75.52	-1.183	0.096	0.98	-0.3	0.30	0.655
2	bevo08_c	28.79	0.935	0.095	0.99	-0.1	0.25	0.322
3	bevo15_c	14.38	1.857	0.118	1.04	0.6	0.05	-0.144
4	bfkt02_c	51.75	-0.075	0.083	0.97	-1.6	0.34	0.733
5	bfkt09_c	45.18	0.202	0.084	1.01	0.3	0.28	0.423
6	bfkt16_c	62.99	-0.559	0.087	0.96	-1.6	0.38	0.847
7	bgen03_c	54.07	-0.178	0.085	1.04	2.2	0.18	0.149
8	bgen10_c	62.07	-0.515	0.087	0.99	-0.3	0.31	0.555
9	bgen17_c	53.87	-0.166	0.084	1.05	2.6	0.16	0.101
10	binf04_c	31.80	0.801	0.089	1.01	0.2	0.23	0.363
11	binf11_c	66.56	-0.727	0.088	0.95	-1.5	0.41	1,064
12	binf18_c	41.79	0.345	0.085	1.02	1.2	0.22	0.238
13	boek05_c	59.94	-0.425	0.086	1.01	0.2	0.26	0.346
14	boek12_c	51.12	-0.048	0.084	0.97	-1.6	0.35	0.718

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
15	boek19_c	89.95	-2.277	0.135	1.00	0.0	0.22	0.751
16	bstw06_c	35.07	0.643	0.088	0.95	-1.6	0.38	0.922
17	bstw13_c	65.28	-0.666	0.088	0.98	-0.7	0.31	0.673
18	bstw20_c	45.51	0.185	0.084	1.02	1.2	0.22	0.309
19	bevo02_c	70.58	-0.918	0.091	1.00	0.1	0.29	0.514
20	bevo09_c	56.91	-0.299	0.086	1.00	0.0	0.29	0.506
21	bevo16_c	34.89	0.652	0.088	0.97	-1.1	0.36	0.735
22	bfkt03_c	80.51	-1.480	0.104	0.97	-0.6	0.33	1.256
23	bfkt10_c	61.05	-0.473	0.085	0.99	-0.5	0.31	0.732
24	bfkt17_c	33.11	0.740	0.090	0.99	-0.2	0.30	0.606
25	bgen04_c	34.24	0.684	0.088	0.99	-0.5	0.29	0.555
26	bgen11_c	24.92	1.159	0.096	0.99	-0.3	0.29	0.608
27	bgen18_c	17.64	1.613	0.109	0.99	-0.2	0.26	0.692
28	binf05_c	53.33	-0.140	0.084	1.00	-0.1	0.29	0.456
29	binf12_c	45.87	0.173	0.084	1.03	1.5	0.21	0.240
30	binf19_c	38.65	0.483	0.087	0.96	-1.5	0.38	0.858
31	boek06_c	31.31	0.822	0.092	1.02	0.5	0.21	0.264
32	boek13_c	37.75	0.527	0.087	0.97	-1.0	0.32	0.633

	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	boek20_c	54.96	-0.211	0.085	1.00	-0.1	0.30	0.537
34	bstw07_c	42.60	0.314	0.085	0.98	-0.8	0.31	0.601
35	bstw14_c	70.34	-0.910	0.092	1.01	0.2	0.23	0.380
36	bstw21s_c	n.a.	3.395	0.309	0.97	-0.3	0.32	0.791
37	bevo03_c	53.26	-0.145	0.083	1.07	3.7	0.12	-0.146
38	bevo10_c	36.98	0.554	0.087	1.01	0.4	0.22	0.215
39	bevo17_c	43.98	0.244	0.084	1.03	1.5	0.20	0.153
40	bfkt04_c	53.75	-0.160	0.084	0.97	-1.6	0.35	0.710
41	bfkt11_c	25.81	1.100	0.095	1.06	1.4	0.03	-0.239
42	bfkt18_c	61.49	-0.496	0.085	0.99	-0.4	0.29	0.563
43	bgen05_c	27.56	1.009	0.093	0.95	-1.2	0.39	1.023
44	bgen12_c	59.30	-0.396	0.085	1.02	1.1	0.21	0.238
45	bgen19_c	28.19	0.971	0.094	1.01	0.2	0.22	0.324
46	binf06_c	59.52	-0.409	0.085	1.02	0.7	0.24	0.308
47	binf13_c	37.64	0.525	0.086	1.01	0.4	0.26	0.369
48	binf20_c	41.79	0.346	0.086	0.99	-0.4	0.29	0.543
49	boek07_c	29.77	0.896	0.091	1.02	0.5	0.18	0.161
50	boek14_c	37.42	0.539	0.088	0.97	-1.2	0.35	0.650

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
51	boek21_c	73.62	-1.051	0.104	1.04	0.8	0.14	0.100
52	bstw01_c	29.98	0.882	0.091	1.03	0.7	0.16	0.202
53	bstw08_c	67.15	-0.757	0.089	0.97	-1.1	0.35	0.819
54	bstw15_c	38.12	0.502	0.086	1.02	0.8	0.23	0.308
55	bevo04_c	13.97	1.895	0.118	1.02	0.3	0.15	0.187
56	bevo11_c	17.95	1.575	0.110	1.02	0.3	0.15	0.291
57	bevo18_c	55.14	-0.220	0.083	1.00	-0.1	0.28	0.491
58	bfkt05_c	28.44	0.963	0.091	1.01	0.2	0.25	0.470
59	bfkt12_c	73.57	-1.072	0.092	0.99	-0.3	0.31	0.660
60	bfkt19_c	70.63	-0.920	0.090	0.96	-1.2	0.39	1.158
61	bgen06_c	47.48	0.102	0.083	1.02	1.3	0.22	0.284
62	bgen13_c	28.62	0.957	0.091	1.02	0.4	0.22	0.292
63	bgen20_c	23.00	1.262	0.097	1.02	0.5	0.17	0.212
64	binf07_c	82.66	-1.629	0.107	0.97	-0.5	0.33	0.947
65	binf14_c	56.28	-0.266	0.083	1.01	0.7	0.24	0.281
66	binf21s_c	n.a.	1.879	0.174	0.95	-0.7	0.40	0.907
67	boek01_c	21.47	1.354	0.100	1.02	0.4	0.17	0.129
68	boek08_c	43.20	0.286	0.084	1.02	1.2	0.22	0.215

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
69	boek15_c	36.96	0.558	0.085	1.00	0.2	0.26	0.321
70	bstw02_c	97.08	-3.603	0.234	0.98	-0.0	0.21	1.873
71	bstw09_c	88.32	-2.103	0.125	0.97	-0.3	0.32	1.461
72	bstw16_c	67.85	-0.783	0.088	0.95	-1.6	0.41	1.386
73	bevo05_c	69.05	-0.835	0.089	0.99	-0.4	0.31	0.461
74	bevo12_c	9.59	2.343	0.137	1.05	0.5	-0.03	-0.405
75	bevo19_c	7.14	2.665	0.157	1.01	0.1	0.12	0.231
76	bfkt06_c	43.94	0.264	0.083	0.98	-1.2	0.35	0.602
77	bfkt13_c	29.13	0.940	0.091	1.00	-0.1	0.27	0.517
78	bfkt20_c	77.18	-1.272	0.096	1.02	0.4	0.17	0.267
79	bgen07_c	44.63	0.232	0.083	1.03	1.4	0.21	0.223
80	bgen14_c	70.83	-0.925	0.090	0.97	-0.8	0.35	0.754
81	bgen21s_c	n.a.	0.981	0.163	0.98	-0.3	0.27	0.441
82	binf01_c	34.98	0.658	0.087	1.00	0.2	0.27	0.340
83	binf08_c	41.37	0.374	0.084	0.97	-1.3	0.35	0.660
84	binf15_c	58.35	-0.350	0.084	1.01	0.5	0.25	0.322
85	boek02_c	77.54	-1.291	0.098	0.96	-0.9	0.35	1.065
86	boek09_c	52.34	-0.092	0.083	0.96	-2.1	0.38	0.820

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
87	boek16_c	42.06	0.343	0.084	1.00	-0.1	0.30	0.549
88	bstw03_c	37.54	0.543	0.085	1.06	2.3	0.11	0.027
89	bstw10_c	24.72	1.170	0.096	1.03	0.7	0.14	0.072
90	bstw17_c	63.47	-0.574	0.085	0.95	-2.0	0.41	1.005
91	bevo06_c	30.37	0.862	0.092	1.05	1.3	0.12	0.015
92	bevo13_c	23.95	1.203	0.097	1.02	0.3	0.18	0.301
93	bevo20_c	37.85	0.518	0.087	1.00	-0.1	0.27	0.530
94	bfkt07_c	31.23	0.825	0.089	0.99	-0.2	0.30	0.572
95	bfkt14_c	80.79	-1.492	0.106	0.97	-0.4	0.30	0.853
96	bfkt21_c	50.20	0.019	0.092	0.99	-0.4	0.31	0.632
97	bgen01_c	49.20	0.025	0.084	1.03	1.7	0.21	0.190
98	bgen08_c	46.54	0.138	0.084	0.98	-1.1	0.33	0.677
99	bgen15_c	78.21	-1.338	0.099	1.02	0.5	0.15	0.117
100	binf02_c	56.85	-0.295	0.084	1.02	1.0	0.22	0.279
101	binf09_c	30.11	0.875	0.091	1.00	-0.1	0.28	0.553
102	binf16_c	36.64	0.568	0.087	1.04	1.5	0.15	0.036
103	boek03_c	86.36	-1.924	0.117	0.98	-0.2	0.31	1.182
104	boek10_c	64.89	-0.647	0.086	0.97	-0.9	0.36	0.804

	Item	Percentage correct	Difficulty/ location parameter	SE (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
105	boek17_c	23.70	1.219	0.097	0.96	-0.9	0.36	1.077
106	bstw04_c	89.25	-2.201	0.130	0.96	-0.4	0.36	1.623
107	bstw11_c	33.17	0.726	0.088	0.95	-1.6	0.39	0.909
108	bstw18_c	59.22	-0.397	0.084	1.01	0.3	0.25	0.417
109	bevo07_c	34.62	0.672	0.090	0.98	-0.5	0.31	0.635
110	bevo14_c	58.32	-0.349	0.085	1.02	0.9	0.22	0.346
111	bevo21s_c	n.a.	-0.607	0.102	1.02	0.6	0.21	0.170
112	bfkt01_c	25.12	1.148	0.096	1.00	-0.1	0.23	0.422
113	bfkt08_c	41.05	0.383	0.087	0.98	-0.7	0.32	0.631
114	bfkt15_c	27.00	1.043	0.094	1.07	1.7	0.03	-0.182
115	bgen02_c	56.91	-0.287	0.085	1.07	3.3	0.09	-0.140
116	bgen09_c	33.67	0.714	0.090	1.06	1.8	0.09	-0.081
117	bgen16_c	30.36	0.872	0.092	0.99	-0.3	0.28	0.612
118	binf03_c	51.39	-0.053	0.085	1.01	0.7	0.25	0.325
119	binf10_c	72.17	-0.994	0.093	0.98	-0.6	0.32	0.635
120	binf17_c	51.80	-0.074	0.084	0.99	-0.6	0.31	0.557
121	boek04_c	71.80	-0.973	0.093	1.00	-0.0	0.27	0.568
122	boek11_c	54.85	-0.200	0.084	1.04	2.0	0.17	0.117

	Item	Percentage correct	Difficulty/ location parameter	<i>SE</i> (difficulty/ location parameter)	WMNSQ	WMNSQ t-value	Correlation of item score with total score	Discrimination- 2 PL
123	boek18_c	24.28	1.194	0.097	1.02	0.4	0.19	0.267
124	bstw05_c	52.81	-0.113	0.085	1.01	0.5	0.26	0.402
125	bstw12_c	76.61	-1.236	0.098	0.95	-0.9	0.41	1.276
126	bstw19_c	54.21	-0.173	0.084	0.97	-1.9	0.36	0.703

Appendix D: Content Area for each Item

Table S2.

Content Area for each Items

	Item	Content Area		Item	Content Area
1	bevo01_c	Systematics/ evolution	64	binf07_c	Information processing/ characteristics/ immunobiology
2	bevo08_c	Systematics/ evolution	65	binf14_c	Information processing/ characteristics/ immunobiology
3	bevo15_c	Systematics/ evolution	66	binf21s_c	Information processing/ characteristics/ immunobiology
4	bfkt02_c	Cytology/ anatomy/ metabolism	67	boek01_c	Ecology
5	bfkt09_c	Cytology/ anatomy/ metabolism	68	boek08_c	Ecology
6	bfkt16_c	Cytology/ anatomy/ metabolism	69	boek15_c	Ecology
7	bgen03_c	Genetics/ developmental biology	70	bstw02_c	Cytology/ anatomy/ metabolism
8	bgen10_c	Genetics/ developmental biology	71	bstw09_c	Cytology/ anatomy/ metabolism
9	bgen17_c	Genetics/ developmental biology	72	bstw16_c	Cytology/ anatomy/ metabolism
10	binf04_c	Information processing/ characteristics/ immunobiology	73	bevo05_c	Systematics/ evolution
11	binf11_c	Information processing/ characteristics/ immunobiology	74	bevo12_c	Systematics/ evolution
12	binf18_c	Information processing/ characteristics/ immunobiology	75	bevo19_c	Systematics/ evolution
13	boek05_c	Ecology	76	bfkt06_c	Cytology/ anatomy/ metabolism
14	boek12_c	Ecology	77	bfkt13_c	Cytology/ anatomy/ metabolism
15	boek19_c	Ecology	78	bfkt20_c	Cytology/ anatomy/ metabolism
16	bstw06_c	Cytology/ anatomy/ metabolism	79	bgen07_c	Genetics/ developmental biology

Table S2.
Content Area for each Items

Item		Content Area	Item		Content Area
17	bstw13_c	Cytology/ anatomy/ me-tabolism	80	bgen14_c	Genetics/ developmental biology
18	bstw20_c	Cytology/ anatomy/ me-tabolism	81	bgen21s_c	Genetics/ developmental biology
19	bevo02_c	Systematics/ evolution	82	binf01_c	Information processing/ characteristics/ immunobiology
20	bevo09_c	Systematics/ evolution	83	binf08_c	Information processing/ characteristics/ immunobiology
21	bevo16_c	Systematics/ evolution	84	binf15_c	Information processing/ characteristics/ immunobiology
22	bfkt03_c	Cytology/ anatomy/ metabolism	85	boek02_c	Ecology
23	bfkt10_c	Cytology/ anatomy/ metabolism	86	boek09_c	Ecology
24	bfkt17_c	Cytology/ anatomy/ metabolism	87	boek16_c	Ecology
25	bgen04_c	Genetics/ developmental biology	88	bstw03_c	Cytology/ anatomy/ me-tabolism
26	bgen11_c	Genetics/ developmental biology	89	bstw10_c	Cytology/ anatomy/ me-tabolism
27	bgen18_c	Genetics/ developmental biology	90	bstw17_c	Cytology/ anatomy/ me-tabolism
28	binf05_c	Information processing/ characteristics/ immunobiology	91	bevo06_c	Systematics/ evolution
29	binf12_c	Information processing/ characteristics/ immunobiology	92	bevo13_c	Systematics/ evolution
30	binf19_c	Information processing/ characteristics/ immunobiology	93	bevo20_c	Systematics/ evolution
31	boek06_c	Ecology	94	bfkt07_c	Cytology/ anatomy/ metabolism
32	boek13_c	Ecology	95	bfkt14_c	Cytology/ anatomy/ metabolism
33	boek20_c	Ecology	96	bfkt21_c	Cytology/ anatomy/ metabolism
34	bstw07_c	Cytology/ anatomy/ me-tabolism	97	bgen01_c	Genetics/ developmental biology

Table S2.
Content Area for each Items

Item		Content Area	Item		Content Area
35	bstw14_c	Cytology/ anatomy/ me-tabolism	98	bgen08_c	Genetics/ developmental biology
36	bstw21s_c	Cytology/ anatomy/ me-tabolism	99	bgen15_c	Genetics/ developmental biology
37	bevo03_c	Systematics/ evolution	100	binf02_c	Information processing/ characteristics/ immunobiology
38	bevo10_c	Systematics/ evolution	101	binf09_c	Information processing/ characteristics/ immunobiology
39	bevo17_c	Systematics/ evolution	102	binf16_c	Information processing/ characteristics/ immunobiology
40	bfkt04_c	Cytology/ anatomy/ metabolism	103	boek03_c	Ecology
41	bfkt11_c	Cytology/ anatomy/ metabolism	104	boek10_c	Ecology
42	bfkt18_c	Cytology/ anatomy/ metabolism	105	boek17_c	Ecology
43	bgen05_c	Genetics/ developmental biology	106	bstw04_c	Cytology/ anatomy/ me-tabolism
44	bgen12_c	Genetics/ developmental biology	107	bstw11_c	Cytology/ anatomy/ me-tabolism
45	bgen19_c	Genetics/ developmental biology	108	bstw18_c	Cytology/ anatomy/ me-tabolism
46	binf06_c	Information processing/ characteristics/ immunobiology	109	bevo07_c	Systematics/ evolution
47	binf13_c	Information processing/ characteristics/ immunobiology	110	bevo14_c	Systematics/ evolution
48	binf20_c	Information processing/ characteristics/ immunobiology	111	bevo21s_c	Systematics/ evolution
49	boek07_c	Ecology	112	bfkt01_c	Cytology/ anatomy/ metabolism
50	boek14_c	Ecology	113	bfkt08_c	Cytology/ anatomy/ metabolism
51	boek21_c	Ecology	114	bfkt15_c	Cytology/ anatomy/ metabolism
52	bstw01_c	Cytology/ anatomy/ me-tabolism	115	bgen02_c	Genetics/ developmental biology

Table S2.
Content Area for each Items

Item		Content Area	Item		Content Area
53	bstw08_c	Cytology/ anatomy/ me-tabolism	116	bgen09_c	Genetics/ developmental biology
54	bstw15_c	Cytology/ anatomy/ me-tabolism	117	bgen16_c	Genetics/ developmental biology
55	bevo04_c	Systematics/ evolution	118	binf03_c	Information processing/ characteristics/ immunobiology
56	bevo11_c	Systematics/ evolution	119	binf10_c	Information processing/ characteristics/ immunobiology
57	bevo18_c	Systematics/ evolution	120	binf17_c	Information processing/ characteristics/ immunobiology
58	bfkt05_c	Cytology/ anatomy/ metabolism	121	boek04_c	Ecology
59	bfkt12_c	Cytology/ anatomy/ metabolism	122	boek11_c	Ecology
60	bfkt19_c	Cytology/ anatomy/ metabolism	123	boek18_c	Ecology
61	bgen06_c	Genetics/ developmental biology	124	bstw05_c	Cytology/ anatomy/ me-tabolism
62	bgen13_c	Genetics/ developmental biology	125	bstw12_c	Cytology/ anatomy/ me-tabolism
63	bgen20_c	Genetics/ developmental biology	126	bstw19_c	Cytology/ anatomy/ me-tabolism