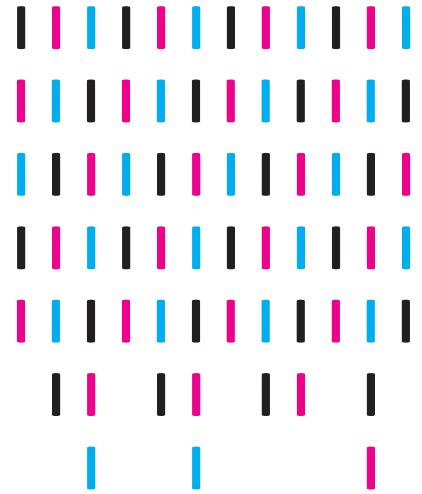




NEPS *SURVEY PAPERS*

Timo Gnambs

NEPS TECHNICAL REPORT FOR
GENERAL KNOWLEDGE:
SCALING RESULTS OF STARTING
COHORT 8 FOR WAVE 2 IN
SPECIAL SCHOOLS

NEPS *Survey Paper* No. 135
Bamberg, June 2026

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).

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 edited by a review board consisting of the scientific management of LIfBi and NEPS.

The NEPS *Survey Papers* are available at www.neps-data.de (see section "Publications") and at www.lifbi.de/publications.

Editor-in-Chief: Thomas Bäumer, LIfBi

Review Board: Board of Directors, Heads of LIfBi Departments, and Scientific Management of NEPS Working Units

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



This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

NEPS Technical Report for General Knowledge: Scaling Results of Starting Cohort 8 for Wave 2 in Special Schools

Timo Gnambs 

Leibniz Institute for Educational Trajectories

E-mail address of the lead author:

timo.gnambs@lifbi.de

Bibliographic data:

Gnambs, T. (2026). *NEPS technical report for general knowledge: scaling results of starting cohort 8 for wave 2 in special schools* (NEPS Survey Paper No. 135). Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/NEPS:SP135:1.0>

Acknowledgements. This report is based on *NEPS Survey Paper No. 134* (Gnambs, 2026) that presents the scaling results for general knowledge in Grade 6 of Starting Cohort 8. Therefore, various parts of this report (e.g., regarding the introduction and the analytic strategy) are reproduced verbatim to facilitate the understanding of the presented results.

NEPS Technical Report for General Knowledge: Scaling Results of Starting Cohort 8 for Wave 2 in Special Schools

Abstract

The National Educational Panel Study (NEPS) examines the development of competencies across the lifespan. Therefore, the NEPS assesses different domains of competence in different age cohorts. To evaluate the quality of these tests, several analyses based on item response theory are conducted. This paper describes the data and scaling procedures for a test of general knowledge that was administered in Grade 6 of Starting Cohort 8 in special schools. The general knowledge test consisted of 64 items representing different content domains. The test was administered to 163 participants (36% girls). A Rasch model was used to scale the data. Item fit statistics, Rasch-homogeneity, and local item independence were evaluated to ensure the quality of the test. The analyses showed that the items had a satisfactory fit to the item response model. In addition, few items showed noteworthy differential item functioning with regard to other school types. However, most of the items were too difficult for the sample. Therefore, the test showed a limited variance and, consequently, a poor reliability. Overall, the test had good psychometric properties but was unable to capture pronounced individual differences in general knowledge of the students. In addition to the scaling results, this report also describes the data available in the scientific use file.

Keywords

item response theory, scaling, general knowledge, scientific use file

Contents

1	Introduction	4
2	Testing General Knowledge	4
2.1	Conceptual Framework	4
2.2	Study Design	6
3	Sample	6
4	Psychometric Analyses	6
4.1	Missing Responses	6
4.2	Scaling Model	6
4.3	Software	7
5	Results	7
5.1	Missing Responses	7
5.1.1	Missing Responses per Person	7
5.1.2	Missing Responses per Item	8
5.2	Parameter Estimates	13
5.2.1	Item Parameters	13
5.2.2	Test Targeting and Reliability	15
5.3	Quality of the test	16
5.3.1	Distractor Analyses	16
5.3.2	Item Fit	16
5.3.3	Differential Item Functioning	17
5.3.4	Rasch-homogeneity	20
5.3.5	Dimensionality	20
6	Discussion	21
7	Data in the Scientific Use Files	21
7.1	Naming Conventions	21
7.2	General Knowledge Scores	21
	References	22
	Appendix A: Content Areas and Domains	23

1. Introduction

The National Educational Panel Study (NEPS)¹ measures different competencies coherently across the lifespan. These include, among others, reading competence, mathematical competence, and domain-general cognitive functioning. An overview of the competencies measured in the NEPS is given by Weinert et al. (2019) and Fuß et al. (2025). Most of the administered competence tests are developed specifically for use in the NEPS and are therefore routinely evaluated using psychometric models based on item response theory (IRT, see Pohl & Carstensen, 2012).

This report presents the results of these analyses for a test of general knowledge that was administered in sixth grades of special schools in Starting Cohort 8 (Grade 5). The following sections first introduce the main concepts of the administered test and the test design. Then, the data are described and the results of various psychometric analyses are reported that were conducted to estimate knowledge scores and to check the quality of the test. Finally, an overview of the data that are available for public use in the scientific use file (SUF) is given.

The analyses summarized in this report are based on the data available at some point prior to the public data release. Due to ongoing data protection and data cleansing issues, the data in the SUF may differ slightly from the data used for the analyses in this report. However, pronounced differences in the presented results are not expected for the final data available in the SUF.

2. Testing General Knowledge

2.1 Conceptual Framework

According to the Cattell-Horn-Carroll (CHC) model of intelligence (Schneider & McGrew, 2018), human cognitive abilities can be described by a hierarchical structure that classifies abilities into different strata. While a single general ability is at the apex of the hierarchy (Stratum III), several broad abilities such as fluid and crystallized intelligence, together with various additional abilities (e.g., memory, cognitive speed), make up Stratum II. In contrast to fluid intelligence, which involves basic reasoning and problem-solving abilities, crystallized intelligence includes acquired skills and knowledge and thus largely reflects learning, education, and acculturation.

Crystallized intelligence can be measured with items that capture the depth and breadth of the dominant knowledge in a culture (Horn & Noll, 1997). Accordingly, the Berlin test of fluid and crystallized intelligence (BEFKI, Schroeders et al., 2020), which was administered in the present study, measures crystallized intelligence with items on factual knowledge in 16 content areas, each with four items (see Table 1). The content areas can be further grouped into three major domains, that is, humanities, natural sciences, and social sciences. The items cover topics that are part of formal education, but also reflect shared cultural knowledge, that is, they describe what children and adolescents are expected to learn and know in a society. Consequently, the test items cover both formal and non-formal knowledge domains. The items do not refer to specialized knowledge, trivial content, or highly time-dependent information (e.g., current sports results or regional train schedules) but focus on general knowledge that is widely considered important and relevant in a society.

¹<https://www.neps-data.de>

Table 1*Content Areas of General Knowledge Test*

Domain	Content Area	Description
Humanities	Literature	Literary work and text types
	Art	Artistic techniques and famous artists
	Music	Musical instruments and musical styles
	Religion	Concepts of different world religions
	Philosophy	Famous philosophers and questions of philosophy
Natural sciences	Physics	Physical quantities and phenomena
	Chemistry	Organic and inorganic chemistry
	Biology	Biological processes, metabolism, and the human body
	Medicine	Human diseases and their treatment
	Geography	Knowledge about geography and geology
	Technology	Knowledge about technical devices and the Internet
Social sciences	History	Historical events and personalities
	Law	Legal institutions and laws in Germany
	Politics	Political institutions and elections in Germany
	Economics	Terms and participants in an economy
	Finances	Payment transactions and financial terms

Table 2*Number of Items by Domain*

Domain	Number of items
Humanities	20
Natural sciences	24
Social sciences	20
Total number of items	64

The general knowledge test consisted of 64 multiple-choice items. Each item was accompanied by four response options, one of which was a correct solution and the other three acted as distractors (i.e., they were incorrect). The distribution of the items across the three domains of the assessment framework is shown in Table 2. The allocation of each item to the content areas and domains is provided in Appendix A.

2.2 Study Design

The study was conducted in small groups at the respective schools of the students. The tests were presented on computer by trained test administrators. The study assessed different competence domains including general knowledge and digital competence (Kutscher, 2026). The general knowledge test was always presented second after the other test. There was no multi-matrix design with respect to the order of the items within the test. All respondents received the test items in the same order. The testing time was limited to 20 minutes. A comprehensive description of the study design is available on the NEPS website².

3. Sample

A total of 163 participants (36% girls) were administered the general knowledge test. However, 2 participants had less than 16 valid responses to the test items. This suggests that these students did not follow the instructions correctly (see Schroeders et al., 2020). Therefore, these cases were excluded from the psychometric analyses and the analyses presented in this report are based on 161 respondents.

4. Psychometric Analyses

4.1 Missing Responses

Competence data contain different types of missing responses. These are missing responses due to a) omitted items and b) items that test takers did not reach. Omitted items occurred when respondents skipped some items. Due to time constraints or lack of motivation, not all participants completed the test within the allotted time. All missing responses after the last valid response were coded as not reached. Missing responses provide information about how well the test worked (e.g., time limits, understanding of instructions, dealing with different response formats). Therefore, the occurrence of missing responses in the test was evaluated to get an idea of how well respondents coped with the test. Missing responses per item were examined to evaluate how well each of the items functioned.

4.2 Scaling Model

Item and person parameters were estimated using a Rasch (1960) model with Gauss-Hermite quadrature (21 nodes). In these analyses, missing responses were scored as incorrect to align with the scoring scheme proposed by the test authors (Schroeders et al., 2020). In order to ensure appropriate psychometric properties of the test, the quality of the test was examined in several analyses.

The items consisted of one correct response option and several distractors (i.e., incorrect response options). The quality of the distractors within the multiple-choice items was examined to evaluate whether the distractors were predominantly chosen by respondents with a lower ability rather than by those who gave a correct response. We therefore calculated the point-biserial correlation between selecting an incorrect response option for a given item and the total correct score for the remaining items. Negative correlations indicate good distractors,

²<https://www.neps-data.de/Data-Center/Overview-and-Assistance/Plausible-Values>

while correlations between .00 and .05 were considered acceptable and correlations above .05 were considered problematic distractors (Pohl & Carstensen, 2012).

The fit of the items to the Rasch (1960) model was evaluated using the WMNSQ statistic, the respective t -value, and the item characteristic curves (see Pohl & 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 investigated further. Correlations of the item score with the corrected total score greater than .30 were considered as good, those greater than .20 as acceptable, and those less than .20 as problematic. The overall judgment of item fit was based on all fit indicators.

The general knowledge test should measure the same construct for all respondents. If some items favored certain subgroups, measurement invariance would be violated and a comparison of competence scores between these subgroups would be biased. In the present study, differential item functioning (DIF) was investigated for school types. DIF was examined using a multigroup item response model, in which the main effects of the subgroups as well as the differential effects of the subgroups on item difficulty were modeled. Based on experience with preliminary data, we considered absolute differences in estimated difficulties between the subgroups that were greater than 1 logit as very strong DIF, absolute differences between 0.6 and 1 as considerable and worthy of further investigation, differences between 0.4 and 0.6 as small but not severe, and differences less than 0.4 as negligible DIF.

The general knowledge test was scaled using the Rasch (1960) model because it preserves the weighting of the different aspects of the framework as intended by the test developers (Schroeders et al., 2020). Nevertheless, Rasch-homogeneity is an assumption that may not hold for empirical data. To test the assumption of equal item discrimination parameters, a two-parametric item response model (2PL, Birnbaum, 1968) was also fitted to the data and compared to the Rasch model.

The dimensionality of the test was evaluated by examining the residuals of the Rasch (1960) model. Approximately zero-order correlations, as indicated by Yen's (1984) Q_3 , suggest unidimensionality. Because the Q_3 tends to be slightly negative in case of locally independent items, we report the corrected Q_3 , which has an expected value of 0. According to Yen (1993), values of Q_3 falling below .20 indicate essential unidimensionality.

4.3 Software

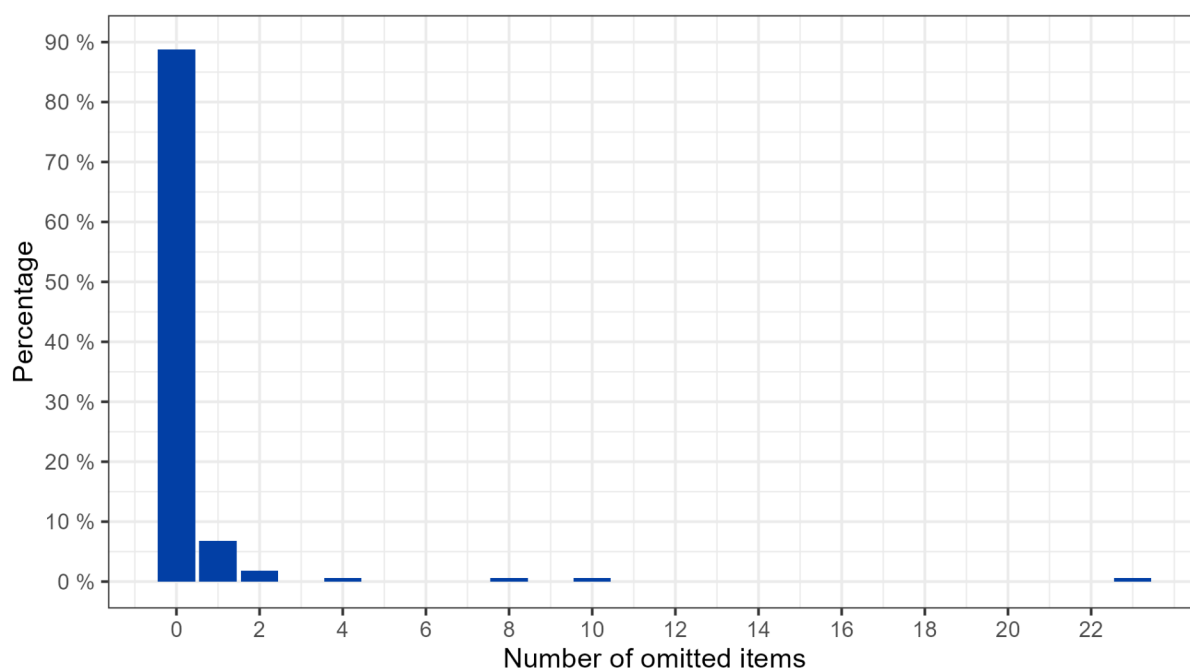
The item response models were estimated with the *TAM* package (Robitzsch et al., 2024) in *R* (R Core Team, 2025).

5. Results

5.1 Missing Responses

5.1.1 Missing Responses per Person

The number of omitted responses per person is shown in Figure 1. The figure shows that only few items were omitted. Approximately 89% of respondents had no omitted items, while 7%

Figure 1*Number of Omitted Items*

exhibited a single omitted item. However, less than 2% of respondents had five or more omitted items.

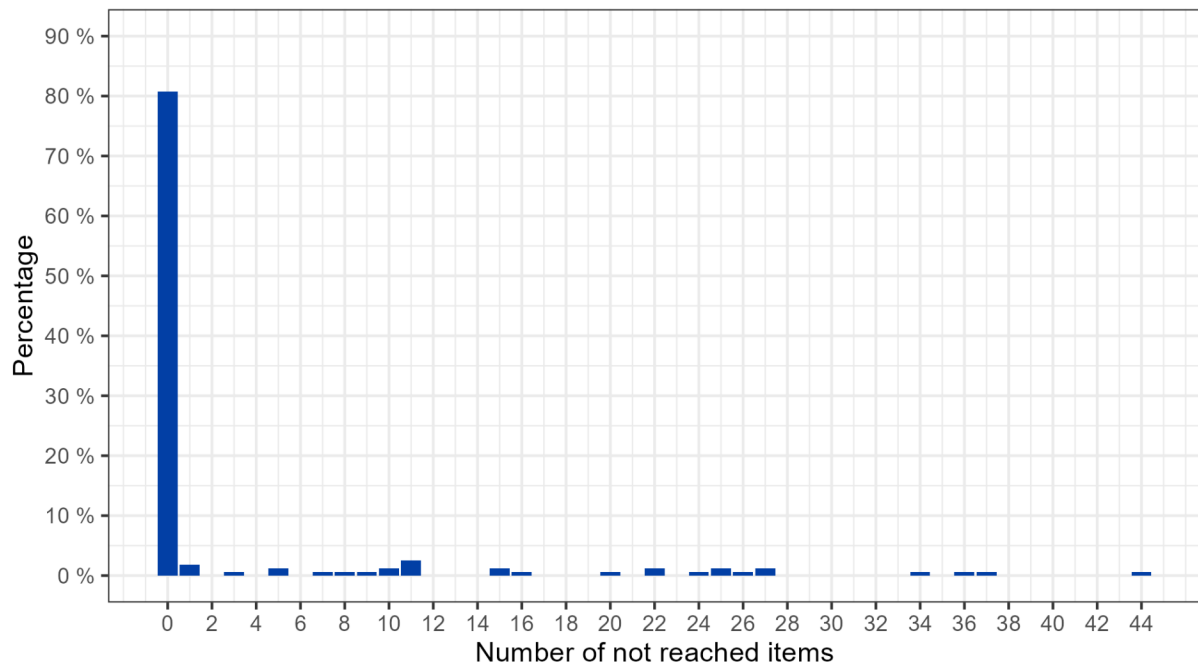
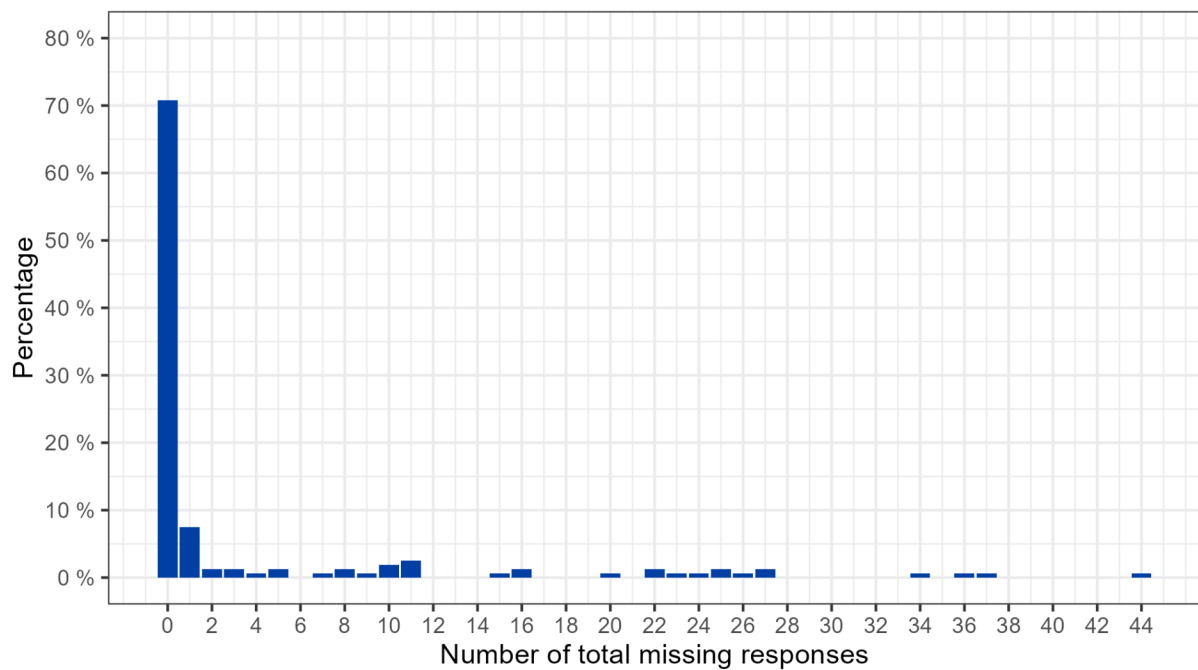
Another source of missing responses is items that respondents did not reach because they aborted the test, for example, due to time constraints or lack of motivation. These missing values refer to items after the last valid response. As shown in Figure 2, about 81% of respondents did not abort the test and were administered all items. Only 16% of the sample received less than 90% of the entire item set.

The total number of missing responses, aggregated across omitted and not reached missing responses for each respondent, is shown in Figure 3. Because most respondents reached the end of the test and had few omitted items, the total number of missing values was small. The median number of missing responses was 0; about 71% of the respondents had no missing response at all. Only 17% of respondents had more than five missing responses, while 10% had missing responses for 16 or more items (i.e., about a quarter of the administered test).

Overall, there was a small amount of omitted and not-reached missing responses. Also, the total number of missing responses was reasonable and did not indicate substantial problems with the test.

5.1.2 Missing Responses per Item

Table 3 provides information on the occurrence of different types of missing responses per item. Overall, the number of respondents that omitted an item was small. The percentage of omitted responses per item varied between 0% and 2% ($Mdn = 1\%$). In addition, most items had relatively few items that were not reached. The percentage of not-reached items varied across items between 0% and 19% ($Mdn = 2\%$). The percentage of respondents not reaching an

Figure 2*Number of Not-Reached Items***Figure 3***Total Number of Missing Responses*

item increased with the position of the item in the test (see Figure 4).

Table 3

Percentage of Missing Responses by Item

Nr.	Item	Pos.	N	OM	NR	ALL
1	gkg6nch1_c	1	160	0.62	0.00	0.62
2	gkg6hph1_c	2	160	0.62	0.00	0.62
3	gkg6nge1_c	3	160	0.62	0.00	0.62
4	gkg6sec1_c	4	158	1.86	0.00	1.86
5	gkg6shi1_c	5	161	0.00	0.00	0.00
6	gkg6nbi1_c	6	161	0.00	0.00	0.00
7	gkg6hli1_c	7	161	0.00	0.00	0.00
8	gkg6nte1_c	8	161	0.00	0.00	0.00
9	gkg6shi2_c	9	160	0.62	0.00	0.62
10	gkg6hmu1_c	10	159	1.24	0.00	1.24
11	gkg6nch2_c	11	160	0.62	0.00	0.62
12	gkg6spo1_c	12	160	0.62	0.00	0.62
13	gkg6shi3_c	13	160	0.62	0.00	0.62
14	gkg6hli2_c	14	160	0.62	0.00	0.62
15	gkg6npy1_c	15	160	0.62	0.00	0.62
16	gkg6hre1_c	16	160	0.62	0.00	0.62
17	gkg6har1_c	17	161	0.00	0.00	0.00
18	gkg6nge2_c	18	161	0.00	0.00	0.00
19	gkg6sla1_c	19	161	0.00	0.00	0.00
20	gkg6nge3_c	20	160	0.62	0.00	0.62
21	gkg6nte2_c	21	159	0.62	0.62	1.24
22	gkg6hli3_c	22	160	0.00	0.62	0.62
23	gkg6nme1_c	23	159	0.62	0.62	1.24
24	gkg6npy2_c	24	159	0.62	0.62	1.24
25	gkg6spo2_c	25	159	0.62	0.62	1.24
26	gkg6hmu2_c	26	160	0.00	0.62	0.62
27	gkg6nch3_c	27	160	0.00	0.62	0.62

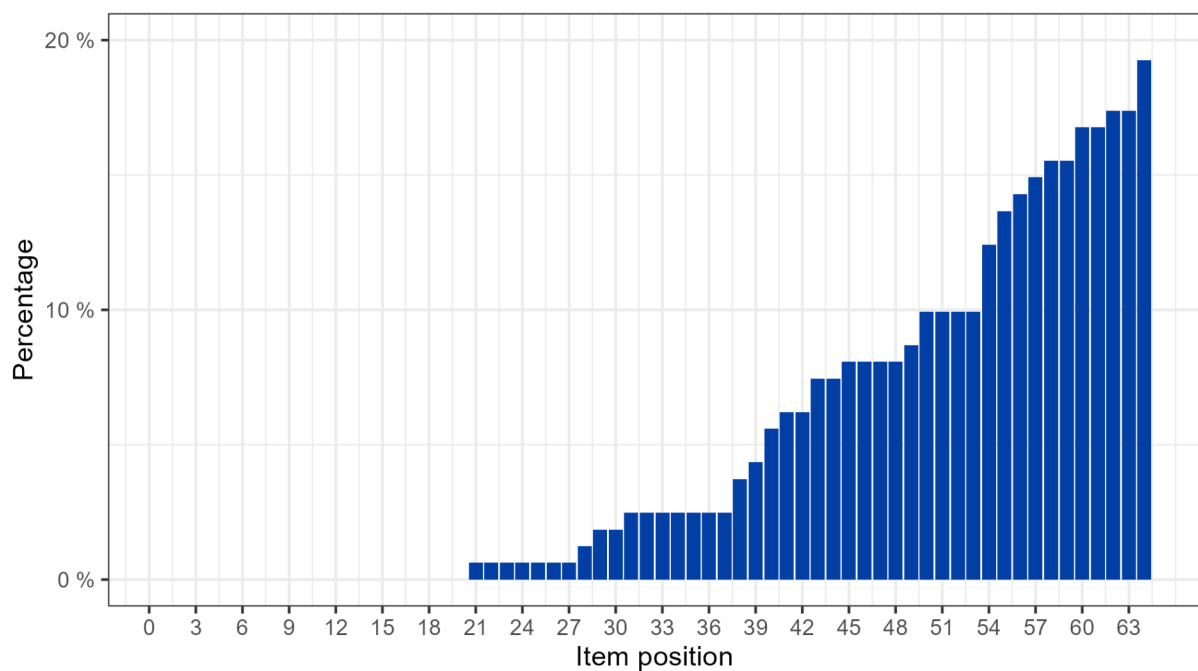
Nr.	Item	Pos.	N	OM	NR	ALL
28	gkg6sla2_c	28	158	0.62	1.24	1.86
29	gkg6hph2_c	29	158	0.00	1.86	1.86
30	gkg6sfi1_c	30	158	0.00	1.86	1.86
31	gkg6nme2_c	31	157	0.00	2.48	2.48
32	gkg6har2_c	32	157	0.00	2.48	2.48
33	gkg6nte3_c	33	156	0.62	2.48	3.11
34	gkg6nch4_c	34	157	0.00	2.48	2.48
35	gkg6hli4_c	35	156	0.62	2.48	3.11
36	gkg6hmu3_c	36	154	1.86	2.48	4.35
37	gkg6sec2_c	37	156	0.62	2.48	3.11
38	gkg6hre2_c	38	151	2.48	3.73	6.21
39	gkg6sfi2_c	39	153	0.62	4.35	4.97
40	gkg6hph3_c	40	152	0.00	5.59	5.59
41	gkg6nbi2_c	41	149	1.24	6.21	7.45
42	gkg6sla3_c	42	150	0.62	6.21	6.83
43	gkg6nme3_c	43	148	0.62	7.45	8.07
44	gkg6spo3_c	44	149	0.00	7.45	7.45
45	gkg6sla4_c	45	147	0.62	8.07	8.70
46	gkg6nme4_c	46	146	1.24	8.07	9.32
47	gkg6npy3_c	47	148	0.00	8.07	8.07
48	gkg6hre3_c	48	145	1.86	8.07	9.94
49	gkg6nge4_c	49	147	0.00	8.70	8.70
50	gkg6hph4_c	50	144	0.62	9.94	10.56
51	gkg6hmu4_c	51	144	0.62	9.94	10.56
52	gkg6sec3_c	52	143	1.24	9.94	11.18
53	gkg6nbi3_c	53	144	0.62	9.94	10.56
54	gkg6sec4_c	54	138	1.86	12.42	14.29
55	gkg6npy4_c	55	135	2.48	13.66	16.15
56	gkg6shi4_c	56	137	0.62	14.29	14.91
57	gkg6hre4_c	57	135	1.24	14.91	16.15

Nr.	Item	Pos.	N	OM	NR	ALL
58	gkg6sfi3_c	58	133	1.86	15.53	17.39
59	gkg6har3_c	59	136	0.00	15.53	15.53
60	gkg6spo4_c	60	134	0.00	16.77	16.77
61	gkg6nte4_c	61	134	0.00	16.77	16.77
62	gkg6sfi4_c	62	133	0.00	17.39	17.39
63	gkg6har4_c	63	131	1.24	17.39	18.63
64	gkg6nbi4_c	64	130	0.00	19.25	19.25

Note. Pos. = Item position within the test. *N* = Number of valid responses. NR = Percentage of respondents that did not reach an item. OM = Percentage of omitted responses. ALL = Percentage of all types of missing responses.

Figure 4

Item Position not Reached



Because a relatively small number of items was omitted or not reached by respondents, the total number of missing responses per item was also quite small. These varied between 0% and 19% (*Mdn* = 2%).

5.2 Parameter Estimates

5.2.1 Item Parameters

The percentage of correct responses out of all valid responses for each item administered varied between 8.07% and 72.67%, thus covering a broad range (see Table 4). The estimated item difficulties are given in Table 4. Item difficulties were estimated by constraining the mean of the ability distribution to zero. The estimated item difficulties ranged from -1.02 (gkg6nge1_c) to 2.47 (gkg6nch2_c) with a median of 0.93. Thus, the test included easy as well as difficult items. Because of the small sample size, the standard errors (*SE*) of the estimated parameters were rather large ($SE \leq 0.29$). Thus, the item difficulties were estimated rather imprecisely.

Table 4

Item Parameters

Nr.	Item	<i>N</i>	Percentage correct	Difficulty	<i>SE</i>	WMNSQ	<i>t</i>	r_{it}	Q_3	Discr.
1	gkg6nch1_c	161	28.57	0.93	0.18	1.03	0.39	0.02	0.07	-0.09
2	gkg6hph1_c	161	30.43	0.84	0.17	0.98	-0.25	0.19	0.08	0.32
3	gkg6nge1_c	161	72.67	-1.02	0.18	0.97	-0.35	0.24	0.07	1.01
4	gkg6sec1_c	161	34.16	0.66	0.17	0.97	-0.54	0.25	0.07	0.74
5	gkg6shi1_c	161	30.43	0.84	0.17	1.01	0.23	0.06	0.07	-0.01
6	gkg6nbi1_c	161	19.25	1.46	0.20	1.02	0.23	0.01	0.06	-0.01
7	gkg6hli1_c	161	21.74	1.30	0.19	1.01	0.12	0.07	0.06	0.35
8	gkg6nte1_c	161	26.71	1.02	0.18	1.00	0.01	0.14	0.07	0.26
9	gkg6shi2_c	161	13.66	1.87	0.23	0.99	0.01	0.10	0.06	0.24
10	gkg6hmu1_c	161	28.57	0.93	0.18	1.04	0.51	-0.05	0.07	-0.18
11	gkg6nch2_c	161	8.07	2.47	0.29	1.02	0.17	-0.05	0.07	-0.17
12	gkg6spo1_c	161	62.11	-0.52	0.16	1.00	0.07	0.11	0.07	0.61
13	gkg6shi3_c	161	22.98	1.23	0.19	1.00	0.02	0.13	0.06	0.45
14	gkg6hli2_c	161	24.22	1.16	0.19	1.03	0.32	0.03	0.05	0.04
15	gkg6npy1_c	161	38.51	0.47	0.16	0.99	-0.20	0.18	0.07	0.71
16	gkg6hre1_c	161	45.96	0.15	0.16	1.00	0.13	0.11	0.06	0.39
17	gkg6har1_c	161	32.92	0.72	0.17	1.01	0.20	0.07	0.06	0.29
18	gkg6nge2_c	161	37.27	0.52	0.17	1.00	0.02	0.14	0.07	0.38
19	gkg6sla1_c	161	34.16	0.66	0.17	1.01	0.13	0.09	0.07	0.15
20	gkg6nge3_c	161	30.43	0.84	0.17	1.01	0.18	0.09	0.06	0.34

Nr.	Item	N	Percentage correct	Difficulty	SE	WMNSQ	t	r_{it}	Q_3	Discr.
21	gkg6nte2_c	161	36.02	0.58	0.17	1.04	0.76	-0.05	0.08	0.04
22	gkg6hli3_c	161	32.92	0.72	0.17	1.03	0.55	0.00	0.08	-0.23
23	gkg6nme1_c	161	48.45	0.05	0.16	0.97	-1.07	0.24	0.06	0.66
24	gkg6npy2_c	161	26.09	1.06	0.18	0.99	-0.06	0.16	0.06	0.29
25	gkg6spo2_c	161	27.33	0.99	0.18	0.96	-0.47	0.29	0.06	0.80
26	gkg6hmu2_c	161	31.68	0.78	0.17	1.03	0.46	0.00	0.06	0.13
27	gkg6nch3_c	161	37.27	0.52	0.17	1.05	1.02	-0.04	0.07	-0.20
28	gkg6sla2_c	161	31.06	0.81	0.17	0.99	-0.08	0.17	0.07	0.42
29	gkg6hph2_c	161	27.33	0.99	0.18	0.98	-0.20	0.22	0.08	0.62
30	gkg6sfi1_c	161	44.10	0.23	0.16	0.98	-0.64	0.21	0.07	0.80
31	gkg6nme2_c	161	28.57	0.93	0.18	1.02	0.30	0.03	0.07	0.00
32	gkg6har2_c	161	18.63	1.50	0.20	1.03	0.30	-0.02	0.06	0.07
33	gkg6nte3_c	161	21.12	1.34	0.20	1.02	0.21	0.04	0.07	0.45
34	gkg6nch4_c	161	30.43	0.84	0.17	0.97	-0.41	0.23	0.06	0.87
35	gkg6hli4_c	161	35.40	0.60	0.17	0.97	-0.51	0.21	0.07	0.38
36	gkg6hmu3_c	161	26.71	1.02	0.18	1.01	0.12	0.11	0.06	0.10
37	gkg6sec2_c	161	32.30	0.75	0.17	1.00	0.09	0.13	0.06	0.19
38	gkg6hre2_c	161	32.30	0.75	0.17	0.97	-0.41	0.21	0.07	0.72
39	gkg6sfi2_c	161	23.60	1.19	0.19	0.99	-0.05	0.13	0.07	0.59
40	gkg6hph3_c	161	39.13	0.44	0.16	1.03	0.64	0.06	0.07	0.09
41	gkg6nbi2_c	161	11.80	2.04	0.25	1.01	0.10	0.05	0.06	0.15
42	gkg6sla3_c	161	21.74	1.30	0.19	0.98	-0.11	0.18	0.08	0.35
43	gkg6nme3_c	161	32.92	0.72	0.17	0.98	-0.26	0.21	0.06	0.51
44	gkg6spo3_c	161	16.15	1.67	0.22	1.01	0.13	0.07	0.06	0.24
45	gkg6sla4_c	161	28.57	0.93	0.18	1.00	-0.01	0.17	0.06	0.22
46	gkg6nme4_c	161	38.51	0.47	0.16	0.99	-0.23	0.19	0.05	0.49
47	gkg6npy3_c	161	34.78	0.63	0.17	0.98	-0.33	0.23	0.08	0.63
48	gkg6hre3_c	161	24.22	1.16	0.19	1.01	0.17	0.09	0.06	0.11
49	gkg6nge4_c	161	22.36	1.26	0.19	1.02	0.20	0.03	0.07	0.10

Nr.	Item	N	Percentage correct	Difficulty	SE	WMNSQ	t	r_{it}	Q_3	Discr.
50	gkg6hph4_c	161	27.95	0.96	0.18	1.00	-0.03	0.13	0.08	0.40
51	gkg6hmu4_c	161	28.57	0.93	0.18	1.00	-0.03	0.14	0.07	0.35
52	gkg6sec3_c	161	19.88	1.42	0.20	1.01	0.16	0.03	0.06	0.01
53	gkg6nbi3_c	161	32.92	0.72	0.17	0.97	-0.50	0.28	0.07	0.81
54	gkg6sec4_c	161	22.98	1.23	0.19	1.05	0.48	-0.06	0.07	-0.03
55	gkg6npy4_c	161	15.53	1.72	0.22	1.00	0.03	0.10	0.06	0.55
56	gkg6shi4_c	161	19.88	1.42	0.20	1.02	0.21	0.03	0.07	0.12
57	gkg6hre4_c	161	22.98	1.23	0.19	1.01	0.11	0.10	0.08	0.24
58	gkg6sfi3_c	161	19.88	1.42	0.20	0.98	-0.13	0.19	0.08	0.61
59	gkg6har3_c	161	28.57	0.93	0.18	0.98	-0.21	0.21	0.06	0.57
60	gkg6spo4_c	161	23.60	1.19	0.19	0.99	-0.11	0.18	0.07	0.60
61	gkg6nte4_c	161	42.24	0.31	0.16	0.92	-2.33	0.45	0.08	2.07
62	gkg6sfi4_c	161	36.02	0.58	0.17	0.96	-0.69	0.26	0.07	0.91
63	gkg6har4_c	161	34.78	0.63	0.17	0.97	-0.53	0.23	0.07	0.49
64	gkg6nbi4_c	161	32.30	0.75	0.17	0.98	-0.27	0.22	0.06	0.42

Note. N = Number of observed responses. Difficulty = Item difficulty. SE = Standard error of item parameter. WMNSQ = Weighted mean square error. t = t -value for WMNSQ. Discr. = Discrimination parameter of a two-parametric item response model. r_{it} = Corrected item-total correlation. Q_3 = Average absolute residual correlation for item (Yen, 1983).

5.2.2 Test Targeting and Reliability

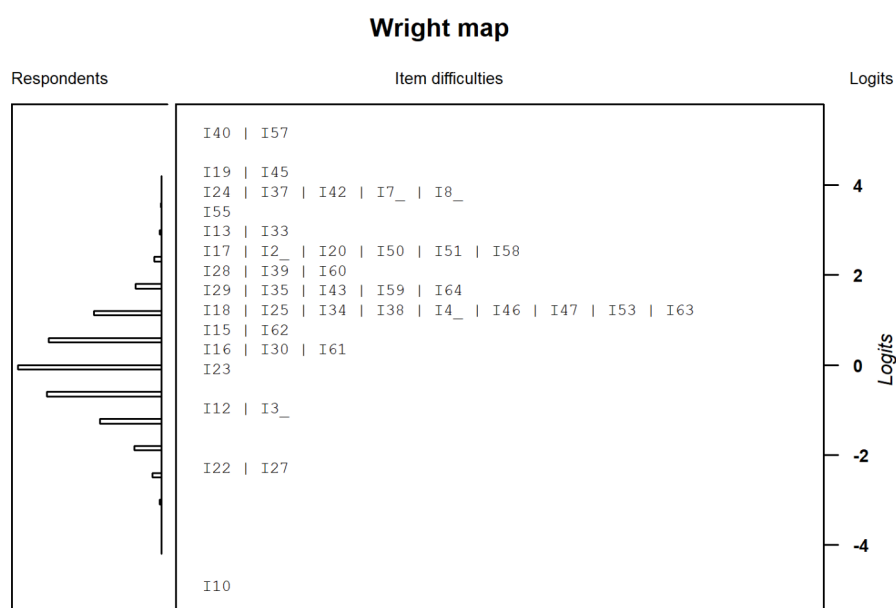
Test targeting focuses on comparing the item difficulties with the person abilities estimated as weighted likelihood estimates (WLE, Warm, 1989) to evaluate the appropriateness of the test for the specific target population. In Figure 5, the item difficulties of the general knowledge items and the ability of the respondents are plotted on the same scale. The distribution of respondents' estimated abilities is plotted on the left, while the distribution of the item difficulties is plotted on the right.

The item difficulties ranged from -1.02 (gkg6nge1_c) to 2.47 (gkg6nch2_c) and thus covered a rather wide range. The mean of the ability distribution was constrained to be zero. The variance was estimated to be 0.12, which implies a rather poor differentiation between respondents. Accordingly, the reliability of the test (EAP/PV reliability = 0.62, WLE reliability = 0.61) was rather low. The median of the item difficulty distribution was about 0.93 logits above the mean person ability distribution. Thus, although the items covered a wide range of the ability distribution, on average, the items were too difficult for the respondents. There were few items covering low and medium ability levels of the sample. As a result, proficiency estimates in higher ability ranges were measured relatively accurately, while low to medium ability estimates had larger

standard errors of measurement.

Figure 5

Test Targeting



Note. The distribution of the person abilities in the sample is given on the left-hand side of the graph. The item difficulties are given on the right-hand side of the graph, with each number corresponding to the item numbers given in Table 4.

5.3 Quality of the test

5.3.1 Distractor Analyses

To investigate how well the distractors of the multiple-choice items performed, point-biserial correlations were calculated between each incorrect response (distractor) and the respondents' total correct scores for the remaining items. The median point-biserial correlation for the distractors fell at $-.06$ ($Min = -.33$, $Max = .26$). Out of a total of 192 distractors, 43 (about 22%) showed correlations greater than $.05$. Overall, these results suggest that most of the distractors worked well. However some distractors appeared to be inappropriate for students in special schools.

5.3.2 Item Fit

The correlations of the correct responses with the total-rest scores varied between -0.06 (gkg6sec4_c) and 0.45 (gkg6nte4_c), with a median of 0.13 . For about a third of the items, the point-biserial correlations were rather small and fell below $.20$; 28 of them had correlations smaller than $.10$.

The evaluation of item fit was based on the scaling model presented above. Overall, the item fit was good (see Table 4). The WMNSQ values ranged from 0.92 (gkg6nte4_c) to 1.05 (gkg6nch3_c,

gkg6sec4_c), with a median of 1.00. There was no item with a WMSNQ greater than 1.15. Also, no t -value for the WMNSQ exceeded 8. These results show that most of the items exhibited a good fit to the Rasch (1960) model.

5.3.3 Differential Item Functioning

The same general knowledge test was administered in regular schools of Starting Cohort 8 (see Gnambs, 2026). To evaluate whether the test measured the same construct for all respondents, differential item functioning (DIF) was examined across school types. The differences between the estimated item difficulties (on the logit scale) in the different subgroups are summarized in Table 5. For example, the column “special vs. lower” reports the differences in item difficulties between special schools and lower secondary modern schools (“Haupt-/Realschule”); a positive value would indicate that the test was more difficult in special schools, whereas a negative value would indicate that the item was less difficult in special schools than lower secondary modern schools.

The sample included 161 respondents from special schools, 799 from lower secondary modern schools (“Haupt-/Realschule”), 2,355 from upper secondary modern schools (“Gymnasium”), and 996 from upper secondary modern schools without differentiation (“Gesamtschule”). There were substantial differences in general knowledge between students from special schools and the other school types as indicated by the main effects of -0.51 logits (Cohen’s $d = -0.83$), -1.20 logits (Cohen’s $d = -1.96$), and -0.54 logits (Cohen’s $d = -0.88$), respectively. These reflect lower general knowledge in special schools than in other school types.

There were 1, 7, and 1 items with strong DIF exceeding 1 between special schools and lower secondary modern schools (gkg6nch3_c), upper secondary modern schools (gkg6nge1_c, gkg6nte2_c, gkg6nch3_c, gkg6nme2_c, gkg6har2_c, gkg6har3_c, gkg6nte4_c), and secondary modern schools without differentiation (gkg6nch3_c). In addition, considerable DIF between 0.6 and 1.0 logits was observed for 7, 13, and 8 items, respectively. These results suggest that some items functioned differently in special schools, particularly in comparison to upper secondary modern schools.

Despite the observed DIF for several items, this hardly affected the main effects for the comparisons involving special schools. These were similar regardless whether DIF was acknowledged or not (see Table 5).

Table 5
Differential Item Functioning

Item	School type		
	special vs. lower	special vs. mixed	special vs. upper
gkg6nch1_c	-0.14 (-0.23)	-0.16 (-0.26)	-0.13 (-0.21)
gkg6hph1_c	-0.03 (-0.05)	0.06 (0.10)	0.39 (0.64)
gkg6nge1_c	0.93 (1.52)	0.85 (1.39)	1.55 (2.54)
gkg6sec1_c	0.36 (0.59)	0.29 (0.47)	0.55 (0.90)

Item	School type		
	special vs. lower	special vs. mixed	special vs. upper
gkg6shi1_c	-0.18 (-0.29)	-0.11 (-0.18)	-0.10 (-0.16)
gkg6nbi1_c	0.00 (0.00)	0.05 (0.08)	-0.10 (-0.16)
gkg6hli1_c	-0.18 (-0.29)	-0.25 (-0.41)	-0.52 (-0.85)
gkg6nte1_c	0.00 (0.00)	0.07 (0.11)	-0.30 (-0.49)
gkg6shi2_c	0.32 (0.52)	0.19 (0.31)	-0.39 (-0.64)
gkg6hmu1_c	-0.01 (-0.02)	0.01 (0.02)	0.35 (0.57)
gkg6nch2_c	-0.54 (-0.88)	-0.56 (-0.92)	-0.91 (-1.49)
gkg6spo1_c	0.14 (0.23)	0.05 (0.08)	0.48 (0.79)
gkg6shi3_c	0.05 (0.08)	-0.21 (-0.34)	-0.70 (-1.14)
gkg6hli2_c	-0.28 (-0.46)	-0.30 (-0.49)	-0.68 (-1.11)
gkg6npy1_c	0.29 (0.47)	0.27 (0.44)	0.24 (0.39)
gkg6hre1_c	0.13 (0.21)	0.11 (0.18)	0.29 (0.47)
gkg6har1_c	-0.17 (-0.28)	-0.15 (-0.25)	-0.08 (-0.13)
gkg6nge2_c	0.47 (0.77)	0.32 (0.52)	0.41 (0.67)
gkg6sla1_c	-0.38 (-0.62)	-0.33 (-0.54)	-0.39 (-0.64)
gkg6nge3_c	0.07 (0.11)	0.16 (0.26)	0.17 (0.28)
gkg6nte2_c	-0.76 (-1.24)	-0.77 (-1.26)	-1.27 (-2.08)
gkg6hli3_c	0.22 (0.36)	0.31 (0.51)	0.57 (0.93)
gkg6nme1_c	0.16 (0.26)	0.18 (0.29)	-0.18 (-0.29)
gkg6npy2_c	-0.05 (-0.08)	-0.09 (-0.15)	0.23 (0.38)
gkg6spo2_c	-0.22 (-0.36)	-0.07 (-0.11)	0.14 (0.23)
gkg6hmu2_c	-0.27 (-0.44)	-0.08 (-0.13)	-0.18 (-0.29)
gkg6nch3_c	-1.23 (-2.01)	-1.33 (-2.18)	-2.18 (-3.57)
gkg6sla2_c	-0.38 (-0.62)	-0.42 (-0.69)	-0.79 (-1.29)
gkg6hph2_c	-0.20 (-0.33)	-0.24 (-0.39)	-0.36 (-0.59)
gkg6sfi1_c	0.54 (0.88)	0.57 (0.93)	0.86 (1.41)
gkg6nme2_c	-0.55 (-0.90)	-0.51 (-0.83)	-1.14 (-1.86)
gkg6har2_c	-0.91 (-1.49)	-0.97 (-1.59)	-1.22 (-2.00)

Item	School type		
	special vs. lower	special vs. mixed	special vs. upper
gkg6nte3_c	-0.19 (-0.31)	-0.15 (-0.25)	-0.39 (-0.64)
gkg6nch4_c	0.39 (0.64)	0.62 (1.01)	1.00 (1.64)
gkg6hli4_c	0.44 (0.72)	0.70 (1.14)	0.97 (1.59)
gkg6hmu3_c	0.13 (0.21)	0.38 (0.62)	0.10 (0.16)
gkg6sec2_c	-0.68 (-1.11)	-0.72 (-1.18)	-0.97 (-1.59)
gkg6hre2_c	0.18 (0.29)	0.09 (0.15)	0.34 (0.56)
gkg6sfi2_c	0.01 (0.02)	0.34 (0.56)	0.61 (1.00)
gkg6hph3_c	-0.63 (-1.03)	-0.58 (-0.95)	-0.61 (-1.00)
gkg6nbi2_c	0.09 (0.15)	0.05 (0.08)	0.02 (0.03)
gkg6sla3_c	-0.14 (-0.23)	0.00 (0.00)	-0.10 (-0.16)
gkg6nme3_c	0.05 (0.08)	0.02 (0.03)	-0.26 (-0.43)
gkg6spo3_c	0.17 (0.28)	0.10 (0.16)	0.33 (0.54)
gkg6sla4_c	-0.18 (-0.29)	-0.17 (-0.28)	-0.04 (-0.07)
gkg6nme4_c	-0.13 (-0.21)	-0.20 (-0.33)	-0.19 (-0.31)
gkg6npy3_c	-0.18 (-0.29)	-0.08 (-0.13)	-0.36 (-0.59)
gkg6hre3_c	-0.03 (-0.05)	0.04 (0.07)	-0.33 (-0.54)
gkg6nge4_c	-0.24 (-0.39)	-0.28 (-0.46)	-0.47 (-0.77)
gkg6hph4_c	-0.21 (-0.34)	-0.40 (-0.65)	-0.38 (-0.62)
gkg6hmu4_c	0.05 (0.08)	0.12 (0.20)	0.44 (0.72)
gkg6sec3_c	-0.41 (-0.67)	-0.42 (-0.69)	-0.67 (-1.10)
gkg6nbi3_c	0.24 (0.39)	0.24 (0.39)	0.72 (1.18)
gkg6sec4_c	0.27 (0.44)	0.11 (0.18)	0.31 (0.51)
gkg6npy4_c	0.39 (0.64)	0.36 (0.59)	0.55 (0.90)
gkg6shi4_c	0.41 (0.67)	0.30 (0.49)	0.47 (0.77)
gkg6hre4_c	-0.04 (-0.07)	-0.22 (-0.36)	0.00 (0.00)
gkg6sfi3_c	-0.20 (-0.33)	-0.11 (-0.18)	-0.24 (-0.39)
gkg6har3_c	0.84 (1.37)	0.79 (1.29)	1.26 (2.06)
gkg6spo4_c	0.44 (0.72)	0.36 (0.59)	0.12 (0.20)

Item	School type		
	special vs. lower	special vs. mixed	special vs. upper
gkg6nte4_c	0.88 (1.44)	0.85 (1.39)	1.52 (2.49)
gkg6sfi4_c	0.45 (0.74)	0.35 (0.57)	0.74 (1.21)
gkg6har4_c	0.57 (0.93)	0.54 (0.88)	0.96 (1.57)
gkg6nbi4_c	-0.06 (-0.10)	-0.06 (-0.10)	0.06 (0.10)
Main effect (DIF model)	-0.51 (-0.83)	-0.54 (-0.88)	-1.20 (-1.96)
Main effect (Main effect model)	-0.55 (-0.90)	-0.58 (-0.96)	-1.23 (-2.04)

Note. Raw differences between item difficulties (in logits) with standardized differences (Cohen's *d*) in parentheses. special = special schools; lower = lower secondary modern school ("Haupt-/Realschule"); upper = upper secondary modern school ("Gymnasium"); mixed = secondary modern schools without differentiation ("Gesamtschule").

5.3.4 Rasch-homogeneity

An essential assumption of the Rasch (1960) model is that all item discrimination parameters are equal. To test this assumption, a two-parametric item response model (2PL, Birnbaum, 1968) was fitted to the data to estimate the discrimination parameters. The estimated discrimination parameters differed substantially between items (see Table 4). The median discrimination parameter fell at 0.35 (*Min* = -0.23, *Max* = 2.07) and was thus rather low. However, because of the small sample size, the standard errors (*SE*) of the item discriminations parameters were rather large ($SE \leq .46$) and, thus, estimated rather imprecisely

Model fit indices suggested a better fit of the Rasch model (AIC = 12,054, BIC = 12,254, number of parameters = 65) compared to the 2PL model (AIC = 12,060, BIC = 12,454, number of parameters = 128). The Rasch model is also in line with the theoretical concepts underlying the test construction and the scoring approach recommended by the test authors (see Schroeders et al., 2020).

5.3.5 Dimensionality

The dimensionality of the test was investigated by evaluating the correlations between the residuals of the Rasch (1960) model. The adjusted Q_3 statistics (see Table 4) were quite low (*Mdn* = 0.07) - the largest absolute residual correlation was 0.08 (gkg6hph1_c, gkg6nte2_c, gkg6hli3_c, gkg6hph2_c, gkg6sla3_c, gkg6npy3_c, gkg6hph4_c, gkg6hre4_c, gkg6sfi3_c, gkg6nte4_c). Overall, these results indicate an essentially unidimensional test.

6. Discussion

The analyses in the previous sections provided information on the quality of the general knowledge test that was administered in sixth grades of special schools in Starting Cohort 8 of the NEPS. Different types of missing responses were examined, item fit statistics and item characteristic curves were evaluated, and item discriminations were investigated. Further quality inspections were conducted by examining differential item functioning and testing Rasch-homogeneity.

The number of missing values was low suggesting that the test was appropriate for the examined sample and the available testing time. Also, several criteria indicated a good fit of the items to the Rasch (1960) model. However, some items exhibited rather low discriminations. Also, essential unidimensionality could be corroborated. Finally, most items showed no pronounced differential item functioning with regard to other school types, suggesting that comparisons between school types are feasible.

However, most items were too difficult for the students. The test included few items matching the low or medium ability range of the sample. As a consequence, the test was unable to capture pronounced individual differences in general knowledge of the students, resulting in a poor reliability of the test.

In conclusion, the test had satisfactory psychometric properties that allowed the estimation of a unidimensional general knowledge score for the participants. However, the test was not appropriately targeted at the ability level of the sample and therefore exhibited a large measurement error.

7. Data in the Scientific Use Files

7.1 Naming Conventions

The SUF for the starting cohort contains 64 items that are scored dichotomously with 0 indicating an incorrect response and 1 indicating a correct response. Details on the naming conventions of the variables can be found in Fuß et al. (2025).

7.2 General Knowledge Scores

In the SUF, general knowledge scores are provided in the form of sum scores (gkg6_sc3) as recommended by the test authors (Schroeders et al., 2020). Additionally, also domain scores are available that measure students' knowledge in humanities (gkg6_sc3a), natural sciences (gkg6_sc3b), and social sciences (gkg6_sc3c).

References

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Addison-Wesley.
- Fuß, D., Gnambs, T., Lockl, K., Attig, M., & Nusser, L. (2025). *Competence data in NEPS: Overview of measures and variable naming conventions (Starting Cohorts 1 to 8)*. Leibniz Institute for Educational Trajectories.
- Gnambs, T. (2026). *NEPS Technical Report for general knowledge: Scaling results of Starting Cohort 8 in Grade 6* (NEPS Survey Paper 134). Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/NEPS:SP134:1.0>
- Horn, J. L., & Noll, J. (1997). Human cognitive abilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). Guilford Press.
- Kutscher, T. (2026). *NEPS Technical Report for digital competence: Scaling results of Starting Cohort 8 in Grade 6 in special schools* [NEPS Survey Paper]. Leibniz Institute for Educational Trajectories.
- Pohl, S., & Carstensen, C. H. (2012). *NEPS Technical Report – Scaling the data of the competence tests* (NEPS Working Paper 14). University of Bamberg, National Educational Panel Study. <https://doi.org/10.5157/NEPS:WP14:1.0>
- R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.
- Robitzsch, A., Kiefert, T., & Wu, M. (2024). *TAM: Test analysis modules* (Version 4.2-21) [Computer software]. <https://doi.org/10.32614/CRAN.package.TAM>
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 73–163). Guilford Press.
- Schroeders, U., Schipolowski, S., & Wilhelm, O. (2020). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 5. und 7. Jahrgangsstufe* (1st ed.). Hogrefe.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(2), 427–450. <https://doi.org/10.1007/BF02294627>
- Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., Carstensen, C. H., & Lockl, K. (2019). Development of competencies across the life course. In H.-P. Blossfeld & H.-G. Roßbach (Eds.), *Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 57–82). Springer. https://doi.org/10.1007/978-3-658-23162-0_4
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125–145. <https://doi.org/10.1177/014662168400800201>
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187–213. <https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>

Appendix A: Content Areas and Domains

No.	Item	Content area	Domain
1	gkg6nch1_c	Chemistry	Natural sciences
2	gkg6hph1_c	Philosophy	Humanities
3	gkg6nge1_c	Geography	Natural sciences
4	gkg6sec1_c	Economy	Social sciences
5	gkg6shi1_c	History	Social sciences
6	gkg6nbi1_c	Biology	Natural sciences
7	gkg6hli1_c	Literature	Humanities
8	gkg6nte1_c	Technology	Natural sciences
9	gkg6shi2_c	History	Social sciences
10	gkg6hmu1_c	Music	Humanities
11	gkg6nch2_c	Chemistry	Natural sciences
12	gkg6spo1_c	Politics	Social sciences
13	gkg6shi3_c	History	Social sciences
14	gkg6hli2_c	Literature	Humanities
15	gkg6npy1_c	Physics	Natural sciences
16	gkg6hre1_c	Religion	Humanities
17	gkg6har1_c	Art	Humanities
18	gkg6nge2_c	Geography	Natural sciences
19	gkg6sla1_c	Law	Social sciences
20	gkg6nge3_c	Geography	Natural sciences
21	gkg6nte2_c	Technology	Natural sciences
22	gkg6hli3_c	Literature	Humanities
23	gkg6nme1_c	Medicine	Natural sciences
24	gkg6npy2_c	Physics	Natural sciences
25	gkg6spo2_c	Politics	Social sciences
26	gkg6hmu2_c	Music	Humanities
27	gkg6nch3_c	Chemistry	Natural sciences
28	gkg6sla2_c	Law	Social sciences

No.	Item	Content area	Domain
29	gkg6hph2_c	Philosophy	Humanities
30	gkg6sfi1_c	Finances	Social sciences
31	gkg6nme2_c	Medicine	Natural sciences
32	gkg6har2_c	Art	Humanities
33	gkg6nte3_c	Technology	Natural sciences
34	gkg6nch4_c	Chemistry	Natural sciences
35	gkg6hli4_c	Literature	Humanities
36	gkg6hmu3_c	Music	Humanities
37	gkg6sec2_c	Economy	Social sciences
38	gkg6hre2_c	Religion	Humanities
39	gkg6sfi2_c	Finances	Social sciences
40	gkg6hph3_c	Philosophy	Humanities
41	gkg6nbi2_c	Biology	Natural sciences
42	gkg6sla3_c	Law	Social sciences
43	gkg6nme3_c	Medicine	Natural sciences
44	gkg6spo3_c	Politics	Social sciences
45	gkg6sla4_c	Law	Social sciences
46	gkg6nme4_c	Medicine	Natural sciences
47	gkg6npy3_c	Physics	Natural sciences
48	gkg6hre3_c	Religion	Humanities
49	gkg6nge4_c	Geography	Natural sciences
50	gkg6hph4_c	Philosophy	Humanities
51	gkg6hmu4_c	Music	Humanities
52	gkg6sec3_c	Economy	Social sciences
53	gkg6nbi3_c	Biology	Natural sciences
54	gkg6sec4_c	Economy	Social sciences
55	gkg6npy4_c	Physics	Natural sciences
56	gkg6shi4_c	History	Social sciences
57	gkg6hre4_c	Religion	Humanities
58	gkg6sfi3_c	Finances	Social sciences

No.	Item	Content area	Domain
59	gkg6har3_c	Art	Humanities
60	gkg6spo4_c	Politics	Social sciences
61	gkg6nte4_c	Technology	Natural sciences
62	gkg6sfi4_c	Finances	Social sciences
63	gkg6har4_c	Art	Humanities
64	gkg6nbi4_c	Biology	Natural sciences
