NEPS *SURVEY PAPERS*

Anna Hawrot

PSYCHOMETRIC PROPERTIES OF THE "BELIEFS ABOUT THE DETERMINANTS OF STUDENT ACHIEVEMENT" SCALE IN STARTING COHORTS 3 and 4

NEPS *Survey Paper* No. 87
Bamberg, September 2021

LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

NEPS
National Educational Panel Study

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

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

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.

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

Psychometric Properties of the "Beliefs about the Determinants of Student Achievement" Scale in Starting Cohorts 3 and 4

Anna Hawrot
Leibniz Institute for Educational Trajectories, Bamberg, Germany

E-mail address of lead author:

anna.hawrot@lifbi.de

Bibliographic data:

Hawrot, A. (2021). *Psychometric Properties of the "Beliefs about the Determinants of Student Achievement" Scale in Starting Cohorts 3 and 4* (NEPS Survey Paper No. 87). Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://doi.org/10.5157/NEPS:SP87:1.0

Psychometric Properties of the "Beliefs about the Determinants of Student Achievement" Scale in Starting Cohorts 3 and 4

Abstract

This paper presents information on the source, theoretical background, and psychometric properties of the "Beliefs about the determinants of student achievement" scale used Starting Cohort 3 and Starting Cohort 4 to measure teacher convictions about the role of various individual-, family-, and school factors in determining school achievement. We ran an item-level analysis and checked the scale's reliability, internal structure, and measurement invariance. The items' discriminatory power differed between subscales and were moderate for Family-Related Factors, moderate to high for School-Related Factors, and low to moderate for Child-Related Factors. The response distributions of three items were heavily skewed and their response scales restricted. The analyses confirmed the three-dimensional structure of the scale, although several cross-loadings were present in the model for Starting Cohort 3, Wave 3. The scale was partially scalar invariant across Waves 1 and 5 of Starting Cohort 3. The subscales' internal consistencies were low and reached marginally acceptable values only for School-Related Factors. Overall, the scale has mixed psychometric properties.

Keywords

teacher convictions, teacher orientations, psychometric properties

Acknowledgments

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 3 (doi: 10.5157/NEPS:SC3:7.0.1) and Starting Cohort 4 (doi: 10.5157/NEPS:SC4:9.1.1). From 2008 to 2013, NEPS data was 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.

I would like to thank Fenja Schaupp for formatting the paper.

This report includes text reproduced verbatim or with small modifications from previous NEPS Survey Papers (e.g., Hawrot, 2021a, 2021b).

1. Introduction

The National Educational Panel Study aims at tracking the development of various competencies, describing their patterns, and better understanding how they unfold across the lifespan (Blossfeld et al., 2011). To this end, information is gathered about various potential sources of influence, including the home environment, educational institutions, or the workplace. However, all these factors need to be measured in a stage-sensitive way, that is, in a way that is adjusted to the participants' age as well as to their developmental stage and educational path.

Since school plays a major role in the development of competence, its various characteristics are measured in the study. That includes contextual and structural characteristics, the quality of school- and classroom-level processes as well as orientations of various actors (Bäumer et al., 2019).

This paper presents information on the source, theoretical background, and psychometric properties of the "Beliefs about the determinants of student achievement" scale used in Waves 1, 3, and 5 of Starting Cohort 3, and Wave 1 of Starting Cohort 4 to assess teacher beliefs about the causes of student achievement. The goal of the paper is to document the scale and provide data users with basic information necessary to make an informed decision about use of the scale in the analyses based on the NEPS data or in their own research.

2. Scale Description

The scale measures teacher beliefs about the role that various individual-, family-, and school factors play in determining student achievement. On a general level, teacher beliefs, by shaping teacher behavior, affect classroom processes (Bäumer et al., 2019). They also constitute an important aspect of teacher professional competence (Baumert & Kunter, 2013). Therefore, they are expected to indirectly affect student outcomes.

The scale includes nine items divided into 3 three-item subscales for Family-Related, Child-Related and School-Related Factors. Subjects are asked to indicate to what extent each listed factor influences student achievement. They respond using a four-point rating scale, where the response options are labelled as follows: 1 = *very unimportant*, 2 = *rather unimportant*, 3 = *rather important*, 4 = *very important*. The items come from the study entitled „Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter" (Artelt & Mudiappa, 2014; BiKS, *Codebuch Zum Lehrerfragebogen Welle 1 [BiKS-8-14 Grundschule]*, n.d.).

Table 1 contains the item wording and the corresponding variable names used in the Scientific Use Files. The original German-language wording is available on the project's website (www.neps-data.de). The variables can be found in NEPSplorer by selecting the following construct in the thematic search: "Learning environments – Learning opportunities in formal learning environments – Teacher – Factors influencing educational success".

Table 1

Items of the "Beliefs about the Determinants of Student Achievement" Scale

Variable Name	Subscale	What factors, from your own experience, have a major influence on the academic achievement of the students? How important is ...
e22683a	F	a) ... the parents' financial situation?
e22683b	Ch	b) ... the willingness of the students to make an effort?
e22683c	F	c) ... the parents' educational background?
e22683d	F	d) ... the mother's employment?
e22683e	Ch	e) ... the child's talent?
e22683f	Ch	f) ... the child's language proficiency?
e22683g	S	g) ... the method of teaching?
e22683h	S	h) ... the coordination among the teachers teaching in that class?
e22683i	S	i) ... the teaching quality?

Note. Ch = Child-Related Factors; F = Family-Related Factors; S = School-Related Factors.

3. Method

3.1 Data and Sample

This study used data gathered during the National Educational Panel Study (NEPS, Blossfeld et al., 2011) from Starting Cohort 3 (SC3) and Starting Cohort 4 (SC4). In subsequent waves, selected teachers who taught sampled students were surveyed as context persons. Table 2 contains information about the waves of each starting cohort in which the scale was administered. It is supplemented by information which grade students participating in NEPS attended at a given measurement point. Please note that the sample did not include teachers from special schools.

The scale was administered as a part of a larger questionnaire using the standard testing procedure for a wave. Information on the procedures is available in the data manuals (Skopek et al., 2012, 2013) and interviewer manuals[1].

[1] https://www.neps-data.de/Data-Center/Data-and-Documentation/Starting-Cohort-Grade-5/Documentation; https://www.neps-data.de/Data-Center/Data-and-Documentation/Starting-Cohort-Grade-9/Documentation

Table 2

The Scale Administration in Starting Cohorts 3 and 4

	1	**2**	**3**	**4**	**5**
Wave (School Year)	**(2010/11)**	**(2011/12)**	**(2012/13)**	**(2013/14)**	**(2014/15)**
Starting Cohort 3	G5		G7		G9
Starting Cohort 4	G9				

Note. G = grade.

In SC4, the scale was administered once. However, in SC3, the scale was administered twice to all sampled teachers (Waves 1 and 5) and once to first-time interviewees (Wave 3). As a consequence, the same teacher could fill in the scale twice – in Wave 1 and 5, or in Wave 3 and 5. This report uses data from the first administration only, since this is sufficient for the analysis of the quality of the scale. Moreover, it also solves the problem of inconsistencies that appeared in gender and birth date of teachers surveyed in different waves but having the same identification number assigned. These inconsistencies suggested that identification numbers might not be fully consistent across waves, causing difficulties in identifying dependent data. Thanks to using data gathered during the first-time administration only, teacher samples in subsequent waves did not overlap. We used variable *ex80211,* which contains information which questionnaire was administered to a teacher (first-time or panel interviewee questionnaire) to identify and exclude all repeated administrations.

Table 3 presents the sample sizes in all measurement occasions. The samples include teachers who responded to at least one item of the scale; thus, the number of teachers who filled in at least one item in the whole questionnaire may be different.

Table 3

Sample Sizes in the Scale's Administration Time-Points

	1	**2**	**3**	**4**	**5**
Wave (School Year)	**(2010/11)**	**(2011/12)**	**(2012/13)**	**(2013/14)**	**(2014/15)**
Starting Cohort 3	594		511		371
Starting Cohort 4	1473				

Note. The samples include the first administration only.

3.2 Analytical Procedure

In the first step, we analyzed missing response rates per person and per item. Next, we inspected item distributions to identify potential problems with response scales, e.g., range restrictions. We also calculated items' discriminatory power (item-rest correlations).

Third step involved analyzing the internal structure of the scale. It was divided into several sub-steps; first, confirmatory factor analyses (CFA) were performed for each sample separately; then, if the CFA models did not provide an adequate fit to the data, exploratory factor analyses (EFA) were performed.

The next step involved verifying measurement invariance of the most optimal measurement model across the waves of SC3. The samples in subsequent waves included the first administration of the scale to a teacher. Therefore, they were treated as independent groups. However, the samples is SC3 and SC4 were drawn from the same schools. Therefore, the same teachers could participate in both studies. Although their identification numbers should have been the same in both cohorts, we could not assure that because of the above mentioned inconsistencies in gender and birth date of teachers having the same identification number. As a consequence of the difficulties in identifying dependent data, we decided not to test measurement invariance across SC3 and SC4.

We tested measurement invariance within CFA framework using the automatic testing feature of M*plus* (Muthén & Muthén, 1998-2015), that is, the "MODEL = CONFIGURAL METRIC SCALAR" command and the DIFFTEST procedure. We did not use ΔAFIs for comparison purposes because they are not recommended with the WLSMV estimator (Sass et al., 2014), which was used in all of the analyses. We tested factor loadings and thresholds separately (Muthén & Muthén, 1998-2015). This approach is used in studies (e.g., Guay et al., 2015; McLarnon & Carswell, 2013). However, some researchers argue that loadings and thresholds should be freed in tandem because they simultaneously influence the item characteristic curve (Muthén & Muthén, 1998-2015).

The two last steps consisted of conducting reliability analyses and inspecting the factor score distributions.

All the analyses of internal structure were performed with M*plus* 7.4 (Muthén & Muthén, 1998-2015) using delta parameterization and the WLSMV estimator. This estimator is recommended for ordered categorical data, especially when item response distributions are skewed and the number of response categories is small (e.g., Beauducel & Herzberg, 2006; Flora & Curran, 2004). The scales of CFA factors were set by fixing factor variances at one. EFA models used oblique Geomin rotation. All of the models accounted for the non-independence of teachers clustered within schools by adjusting to the standard errors using a sandwich estimator (CLUSTER option).

The model fit was assessed with three commonly used (McDonald & Ho, 2002) fit indices, that is, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI). We assumed that CFI and TLI values not lower than .95, and RMSEA values not higher than .06 indicated a good fit (Hu & Bentler, 1999).

4. Results

4.1 Missing Responses

NEPS datasets include several codes for missing data. In this study two types of missing value occurred: implausible values and unspecific missing values. Both types refer to nonresponse, with implausible values denoting invalid responses, and unspecific missing values denoting nonresponse for which the cause is unknown.

4.1.1 Missing responses per person

Table 4 contains information with the numbers and percentages of respondents with a given number of implausible values, unspecific missing values, and total missing values. The majority of missing values was unspecific. The number of implausible values per person was negligible. It exceeded 1 for only one participant and in each sample up to 0.8% of respondents provided at least one implausible response.

The rates of unspecific missing values per person were higher than the rates of implausible values. The percentage of respondents with at least one unspecific missing value varied between 5.57% in Wave 1 of SC4 and 7.01% in Wave 5 of SC3. The respondents most often omitted one item, and the rate of single-item omissions ranged between 3.77% (Wave 5, SC3) and 4.88% (Wave 1, SC3).

The rates of total missing values per person and the rates of unspecific missing values per person hardly differed because of the low share of implausible values in the total missing values. Thus, the results for the total missing values are not described.

Table 4

Rates of Missing Values per Person

Number of missing values per person	SC3: W1		SC3: W3		SC3: W5		SC4: W1	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
Implausibe missing values								
0	591	99.49	507	99.22	368	99.19	1469	99.73
1	3	0.51	4	0.78	2	0.54	4	0.27
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0.27	0	0
Total	594	100	511	100	371	100	1473	100
>= 1	3	0.51	4	0.78	3	0.81	4	0.27
Unspecific missing values								
0	557	93.77	478	93.54	345	92.99	1391	94.43
1	29	4.88	23	4.50	14	3.77	61	4.14
2	3	0.51	5	0.98	5	1.35	13	0.88
3	3	0.51	1	0.20	2	0.54	4	0.27
4	0	0	0	0	2	0.54	3	0.20
5	2	0.34	1	0.20	1	0.27	1	0.07
6	0	0	0	0	1	0.27	0	0
7	0	0	3	0.59	0	0	0	0
8	0	0	0	0.00	1	0.27	0	0
Total	594	100	511	100	371	100	1473	100
>= 1	37	6.23	33	6.46	26	7.01	82	5.57

Number of missing values per person	SC3: W1		SC3: W3		SC3: W5		SC4: W1	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Total missing values								
0	556	93.60	475	92.95	343	92.45	1388	94.23
1	29	4.88	25	4.89	14	3.77	63	4.28
2	3	0.51	6	1.17	6	1.62	14	0.95
3	4	0.67	1	0.20	2	0.54	4	0.27
4	0	0	0	0	3	0.81	3	0.20
5	2	0.34	1	0.20	1	0.27	1	0.07
6	0	0	0	0	1	0.27	0	0
7	0	0	3	0.59	0	0	0	0
8	0	0	0	0	1	0.27	0	0
Total	594	100	511	100	371	100	1473	100
>= 1	38	6.1	36	7.05	28	7.55	85	5.77

Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

4.1.2 Missing responses per item

Table 5 contains information about implausible, unspecific missing, and total missing values per item. Implausible values occurred incidentally. The rates of unspecific missing values per item ranged between 0% and 4.85% depending on the item and sample. In all samples, item e22683d showed increased rates.

The rates of total missing values per item and the rates of unspecific missing values per item hardly differed because of the low share of implausible values in the total missing values. As a consequence, the rates of total missing values are not described.

In summary, the rates of implausible values per item were negligible. In conjunction with the negligible rates per person, this result suggests that respondents did not experience major difficulties with using the scale's response format. The rates of unspecific missing values per item and per person were acceptable. However, item e22683d showed increased rates of unspecific missing values in all samples compared to the other items. It is not clear why this happened. Possibly, a fraction of respondents did not have a clear opinion on the role of maternal employment or did not want to express one and omitted the item. However, other explanations cannot be excluded.

Table 5

Rates of Missing Values per Item

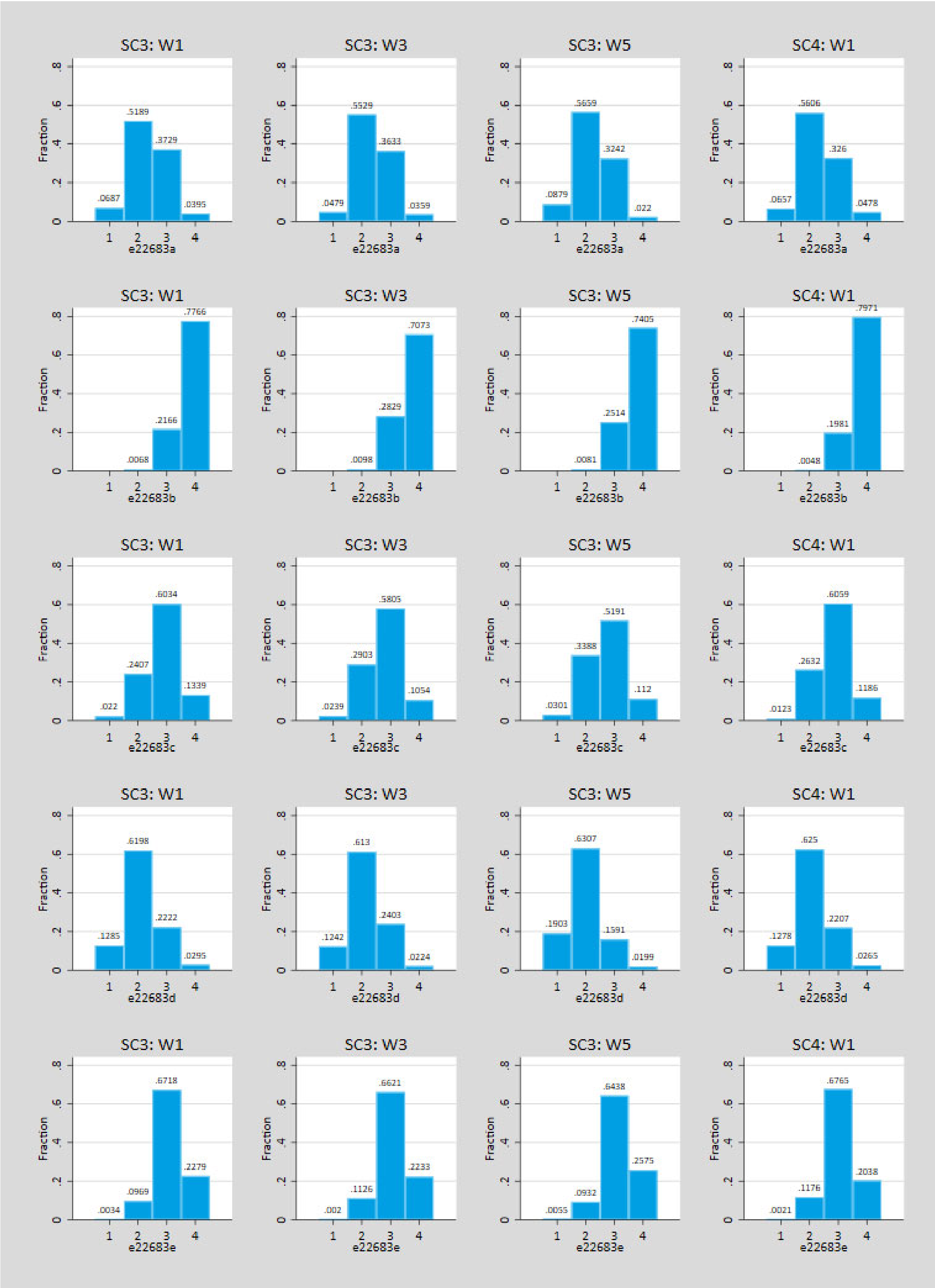
Item	SC3: W1		SC3: W3		SC3: W5		SC4: W1	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
Implausibe value								
e22683a	1	0.17	0	0	0	0	0	0
e22683b	1	0.17	1	0.20	1	0.27	0	0
e22683c	0	0	1	0.20	1	0.27	1	0.07
e22683d	0	0	1	0.20	1	0.27	0	0
e22683e	0	0	0	0	0	0	2	0.14
e22683f	0	0	0	0	1	0.27	1	0.07
e22683g	0	0	0	0	1	0.27	0	0
e22683h	0	0	1	0.20	1	0.27	0	0
e22683i	1	0.17	0	0	0	0	0	0
Unspecific missing								
e22683a	11	1.85	10	1.96	7	1.89	28	1.90
e22683b	2	0.34	1	0.20	0	0	4	0.27
e22683c	4	0.67	7	1.37	4	1.08	13	0.88
e22683d	18	3.03	19	3.72	18	4.85	41	2.78
e22683e	6	1.01	5	0.98	6	1.62	9	0.61
e22683f	2	0.34	3	0.59	5	1.35	3	0.20
e22683g	6	1.01	7	1.37	5	1.35	4	0.27
e22683h	2	0.34	6	1.17	6	1.62	6	0.41
e22683i	3	0.51	4	0.78	6	1.62	8	0.54

Item	SC3: W1		SC3: W3		SC3: W5		SC4: W1	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Total missing								
e22683a	12	2.02	10	1.96	7	1.89	28	1.90
e22683b	3	0.51	2	0.39	1	0.27	4	0.27
e22683c	4	0.67	8	1.57	5	1.35	14	0.95
e22683d	18	3.03	20	3.91	19	5.12	41	2.78
e22683e	6	1.01	5	0.98	6	1.62	11	0.75
e22683f	2	0.34	3	0.59	6	1.62	4	0.27
e22683g	6	1.01	7	1.37	6	1.62	4	0.27
e22683h	2	0.34	7	1.37	7	1.89	6	0.41
e22683i	4	0.67	4	0.78	6	1.62	8	0.54

Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

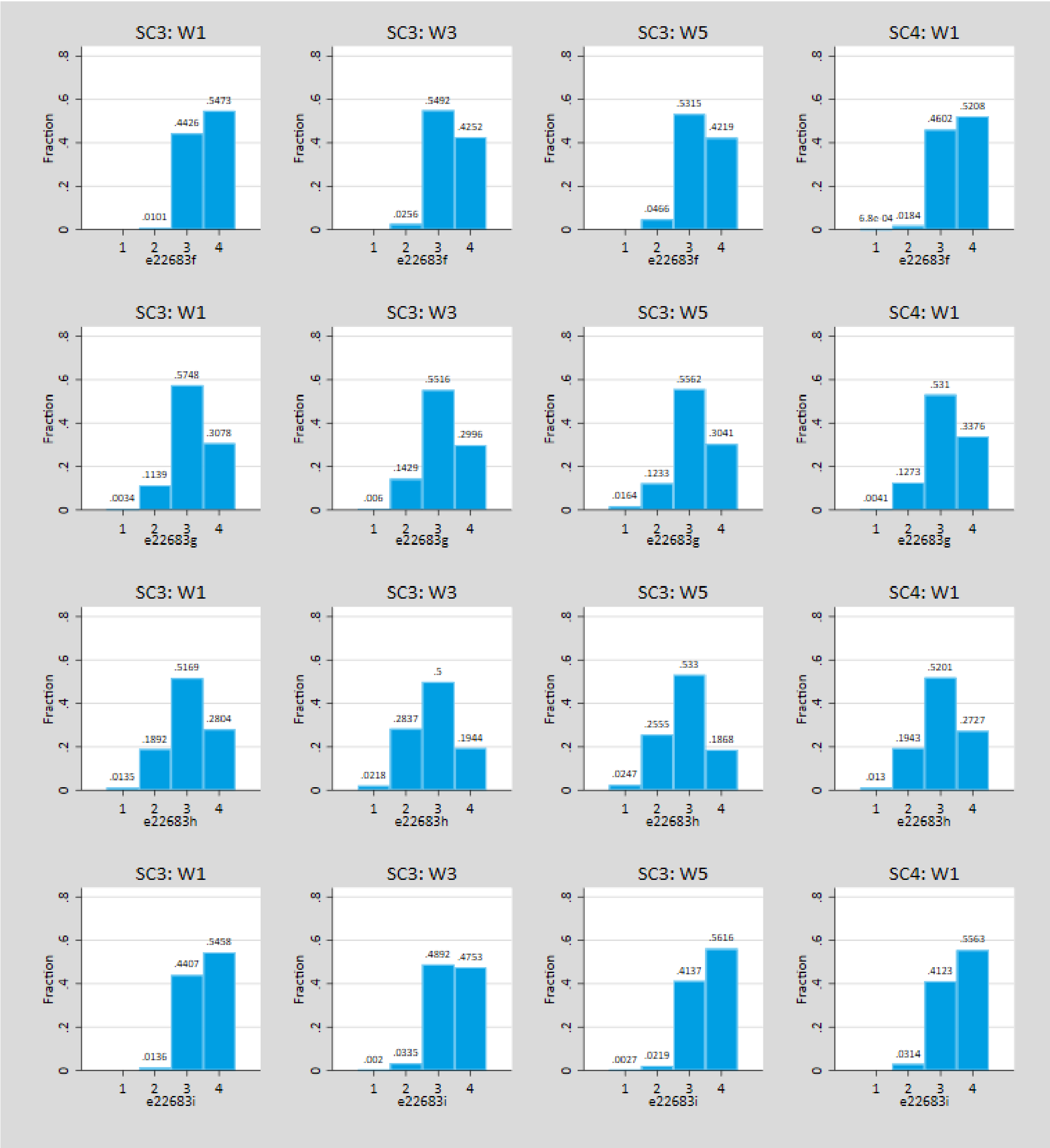
4.2 Item Response Distributions

Figures 1 and 2 present the item response distributions in all samples. Their analysis revealed some unfavorable properties of the items. First, the distributions of several items were skewed or severely skewed; moreover, ceiling effects were visible (see e.g., e22683b, e22683f, e22683i). No responses were recorded for the category *very unimportant* for item e22683b in all samples, for item e22683f in SC3, and for item e22683i in Wave 1 of SC3 and Wave 1 of SC4. This resulted in response scale restrictions. Moreover, for these items low percentages of responses were also recorded for the second lowest category (*relatively unimportant*). Although for the other items some responses were recorded for all of the response categories, the percentages of *very unimportant* category were often very low and did not exceed 0.03% (see e.g., e22683e, e22683g, e22683h).



Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4, W = Wave; 1 = very unimportant, 2 = rather unimportant, 3 = rather important, 4 = very important.

Figure 1. Item response distributions: items e22684a - e22684e.



Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4, W = Wave; 1 = very unimportant, 2 = rather unimportant, 3 = rather important, 4 = very important.

Figure 2. Item response distributions: items e22684f - e22684i.

4.3 Discriminatory Power

To assess the discriminatory power of the items, we calculated item-rest correlations (Spearman's rho) within each subscale. Their values were moderate (.28 to .45) for Family-Related Factors, moderate to high (.38 to .53) for School-Related Factors, but low to moderate (.11 to .39) for Child-Related Factors. The results are presented in Table 6.

Table 6

Item-Rest Correlations (Spearman's rho)

Subscale/Item	**SC3: W1**	**SC3: W3**	**SC3: W5**	**SC4: W1**
Family-Related Factors				
e22683a	.403	.366	.390	.381
e22683c	.450	.347	.386	.360
e22683d	.384	.300	.276	.359
Child-Related Factors				
e22683b	.232	.281	.108	.171
e22683e	.321	.348	.274	.235
e22683f	.319	.390	.220	.263
School-Related Factors				
e22683g	.469	.534	.445	.518
e22683h	.414	.388	.381	.445
e22683i	.473	.467	.425	.501

Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

4.4 Internal Structure

Next, we tested the measure's internal structure. To increase the chances that the sample consisted of respondents who were committed to filling in the scale and provided valid responses, we excluded teachers who had three or more missing values in the scale (over 25%).

In the first step, we ran confirmatory factor analysis for each sample separately to test if the expected three-factor structure held. The models did not include any cross-loadings, but factors were allowed to correlate. All of the models but one (Wave 3 of SC3) showed a good fit to the data. Detailed information is presented in Table 7.

Second, due to a poor fit, we inspected the CFA model for Wave 3 of SC3 and its modification indices. The results suggested that cross-loadings could be a reason of the poor fit, thus we ran a three-factor EFA to check for cross-loadings. The model had a good fit (χ^2 (12, n = 506) = 24.24, p = .019, RMSEA = .045, CFI = .985, TLI = .956). In the model, all items but two (e22683b, e22683d) cross-loaded on another subscale; the cross-loading of item e22683e was the highest and equaled -0.34. Details are presented in Table 8.

Table 7

Fit of the Three-Factor CFA Models

Wave	***N***	**N_{par}**	χ^2	***df***	***p***	**RMSEA**	**CFI**	**TLI**
Starting Cohort 5[a]								
1	588	36	46.83	24	.004	.040	.971	.956
3	506	36	88.63	24	< .001	.073	.922	.883
5	363	36	41.82	24	.014	.045	.956	.934
Starting Cohort 9								
1	1465	37	98.27	24	< .001	.046	.964	.947

Note. N_{par} = number of free parameters.
The Number of free parameters differs between starting cohorts because of response scale restrictions in SC3.
[a] Response categories *very unimportant* (1) and *relatively unimportant* (2) were merged for item e22683i to assure the same number of categories all the waves of SC3.

Nevertheless, in all CFA and EFA models the three expected dimensions, that is Family-Related Factors, Child-Related Factors, and School-Related Factors, appeared. The loadings (and salient loadings in the EFA model for Wave 3 of SC3) varied between the samples, ranging from 0.286 (e22683b, Wave 5 of SC3) to 0.80 (e22683e & e22683i in Wave 1 of SC4, and e22683c in Wave 5 of SC3). Loadings below 0.4 appeared twice only, whereas 31 loading (86.1%) equaled 0.5 or more, and 13 loadings (36.1%) equaled 0.7 or more. The extracted factors showed null to moderate positive correlations in the CFA models, and null correlations in the EFA model. This pattern suggests that cross-loadings might be present in all samples, but in the CFA models, where they were not allowed, they caused inflated factor correlations. However, the good fit of the CFA models suggested that the simple structure was a good approximation of the true factor structure. All factor loadings and factor correlations in the final CFA and EFA solutions are available in Table 8.

In summary, the expected internal structure was confirmed in all waves and starting cohorts. The magnitude of (salient) factor loadings, and cross-loadings in case of the one EFA model, although differentiated, was acceptable.

Table 8

Factor Loadings and Factor Correlations in the Final CFA and EFA Models

Item	SC3: W1[a]			SC3: W3[a]			SC3: W5[a]			SC4: W1		
	Family	**Child**	**School**	**Family[b]**	**Child[b]**	**School[b]**	**Family**	**Child**	**School**	**Family**	**Child**	**School**
e22683a	0.626*			**0.725***	-0.008	0.221*	0.594*			0.617*		
e22683b		0.469*		-0.039	**0.567***	0.098		0.286*			0.349*	
e22683c	0.779*			**0.634***	0.207*	-0.021	0.800*			0.637*		
e22683d	0.613*			**0.549***	0.076	0.025	0.437*			0.585*		
e22683e		0.544*		0.005	**0.686***	-0.337*		0.503*			0.404*	
e22683f		0.719*		0.206*	**0.667***	0.004		0.653*			0.802*	
e22683g			0.737*	0.028	0.210*	**0.742***			0.736*			0.780*
e22683h			0.572*	0.117*	-0.015	**0.591***			0.543*			0.624*
e22683i			0.763*	-0.044	0.244*	**0.686***			0.734*			0.800*
Child	0.165*			-0.035			0.355*			0.410*		
School	0.145*	0.176*		-0.177	0.224		0.059	0.109		0.096*	0.323*	

Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave; Family = Family-Related Factors; Child = Child-Related Factors; School = School-Related Factors.

[a] Response categories *very unimportant* and *relatively unimportant* were merged for the item e22683i to assure the same number of response categories in all the waves of SC3.

[b] Salient factor loadings are in bold type.

* Statistically significant at $p < 0.05$.

4.4.1 Measurement invariance

We checked the measurement invariance across the waves of SC3 that had the same internal structure (Wave 1 and Wave 5). However, the analysis of item response distributions, described in section 4.2, revealed that in some waves no responses were recorded for the category *very unimportant* for several items, which resulted in between-wave differences in the number of response categories. Meanwhile, the analytical methods that we used to test measurement invariance across SC3 required the same number of response categories for a given item in each wave. Thus, if an item had a restricted response scale in at least one wave of SC3, we merged the adjacent categories *very unimportant* and *rather unimportant* in the other waves of SC3 to assure the same number of response categories in all analyses of internal structure.

The test showed that the scale is configurally and metrically but not scalarly invariant. However, it was possible to establish partial scalar invariance by freeing selected thresholds of items e22683f and e22683h. Table 9 presents the results.

Table 9

Results of Testing Measurement Invariance Across Wave 3 and Wave 5 of SC3

model	N_{par}	RMSEA	CFI	TLI	χ^2	*df*	*p*	Comp.	$\Delta\chi^2$ *(df), p*
Starting Cohort 3[a]									
(1)	72	.042	.966	.949	88.55	48	< .001		
(2)	66	.038	.969	.959	90.51	54	.001	2 vs 1	3.36 (6), ns
(3)	54	.043	.951	.947	124.22	66	< .001	3 vs 2	37.86 (12), p < .001
(4)[b]	57	.036	.968	.964	100.88	63	.001	4 vs 2	10.71 (9), ns

Note. (1) = configural; (2) = metric; (3) = scalar; (4) = partial scalar; N_{par} = number of parameters; Comp. = compared models.

[a] N_{Wave1}: 588, N_{Wave5}: 363. Two lowest response categories of item e22683i (*very unimportant* and *relatively unimportant*)), were merged to assure the same number of response categories across the tested waves.

[b] Second and/or third thresholds of the following items were freed: e22683f, e22683h.

4.5 Reliability

In the next step we assessed the reliability of the scale using information on the items' explained variance and total information curves retrieved from the accepted CFA and EFA models. Moreover, we calculated Cronbach's α coefficients based on raw scores. However, the coefficients should be treated with caution because the scale did not meet the assumption of essential tau-equivalence. Factor loadings differed between the items, which means that the items did not measure the latent trait on the same scale. As a consequence, the scale's reliability as measured with Cronbach's α is probably underestimated (Miller, 1995).

Table 10 presents the items' explained variance in models for all samples[2]. They differed between the items and samples, ranging from .082 to .658. A total of 11 out of 36 item-sample combinations had values of .5 or higher. The items belonging to the Child-Related Factors subscale (e22683b, e22683d, e22683e) had the lowest values.

Figure 3 presents the total information curves of the three factors. The measurement precision of the Family-Related Factors subscale was approximately the same at all trait levels. The Child-Related Factors and School-Related Factors subscales showed a decreasing trend at above-mean levels.

Table 10

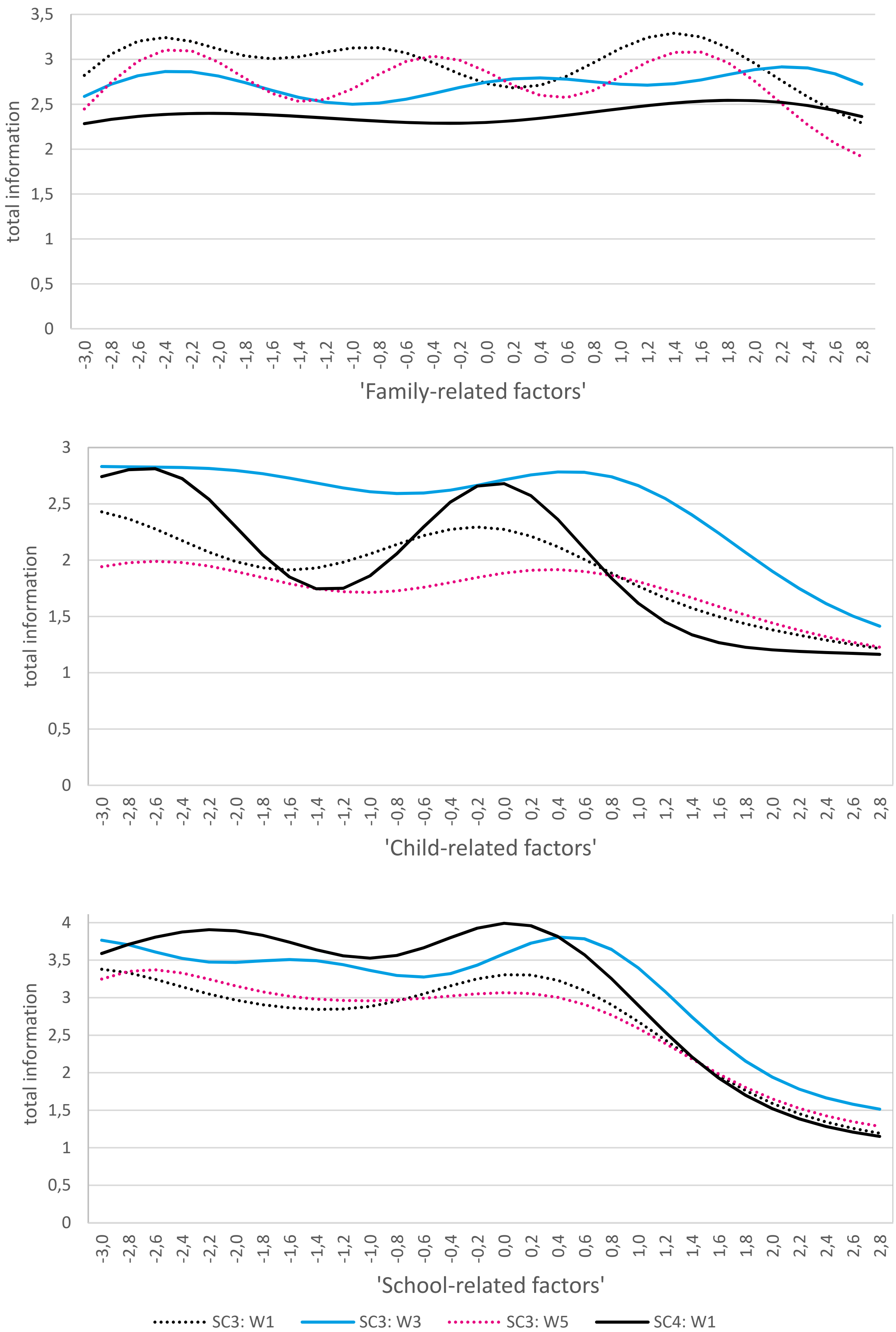
Items' Explained Variance (R^2) in the Final CFA and EFA Models

Item	SC3			SC4
	Wave 1[a]	**Wave 3[b]**	**Wave 5[a]**	**Wave 1[a]**
e22683a	.392	.518	.353	.381
e22683b	.220	.360	.082	.122
e22683c	.606	.438	.640	.405
e22683d	.375	.300	.191	.342
e22683e	.296	.481	.253	.164
e22683f	.516	.479	.426	.643
e22683g	.543	.658	.542	.609
e22683h	.327	.334	.295	.389
e22683i	.582	.619	.539	.640

Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4.
[a] Retrieved from the CFA model for this wave.
[b] Retrieved from the EFA model for this wave.

Table 11 contains information on the internal consistency of the scale. Only the School-Related Factors subscale had marginally acceptable reliability. However, such low values are common in short scales.

[2] Please note that all calculations were performed using the WLSMV estimator and therefore based on the polychoric correlation matrix. As a consequence, the explained variances refer to continuous underlying response variables instead of to categorical observed response variables.



Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

Figure 3. Total information curves of CFA and EFA factors

Table 11

Cronbach's α Coefficients for the Subscales

Subscale	SC3			SC4
	Wave 1	Wave 3	Wave 5	Wave 1
Family-Related Factors	.633	.577	.555	.568
Child-Related Factors	.452	.512	.348	.373
School-Related Factors	.618	.651	.610	.678

Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4.

4.6 Factor Scores Distributions

Figure 4 presents the distributions of the factor scores derived from the final CFA and EFA models. Some deviation from normality was present, including non-symmetry (e.g., Child-Related Factors in Wave 5 of SC4), ceiling effects (e.g., School-Related Factors in Wave 1 of SC3), or multi-modality (e.g., Child-Related Factors in Wave 1 of SC4). Descriptive statistics are available in the Appendix.

5. Summary

This paper documents the "Beliefs about the determinants of student achievement" scale used in Starting Cohort 3 and 4 to assess teacher convictions about the role of various individual-, family-, and school factors in determining school achievement. Besides providing basic information about the scale' source and theoretical background, the paper reports information on its psychometric properties.

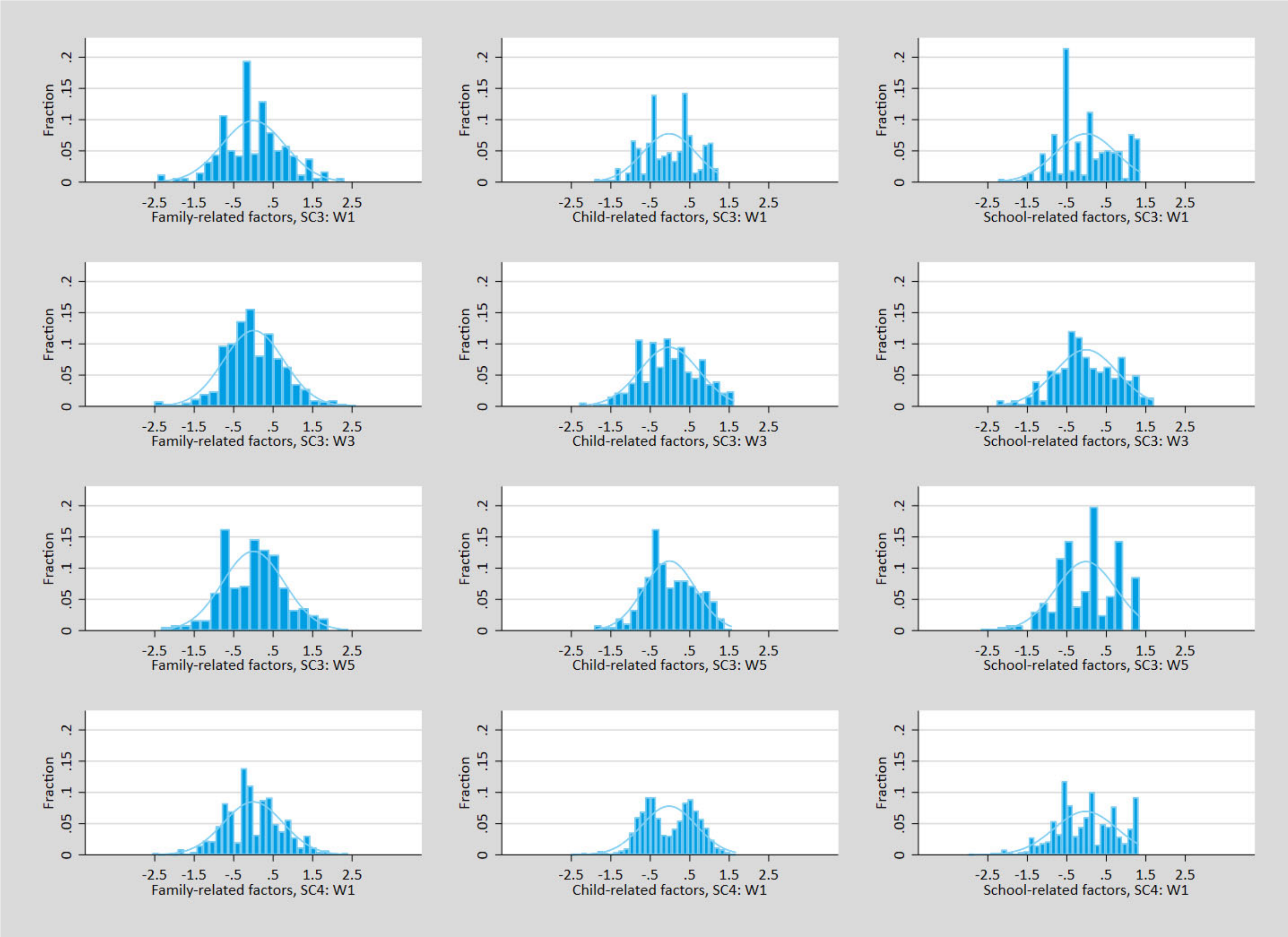
The scale was administered in Waves 1, 3, and 5 of Starting Cohort 3 and Wave 1 of Starting Cohort 4 to teachers who taught sampled students. The analyses included 2,949 respondents in total. They revealed that the rates of missing values per item and per person were satisfactory. Up to 0.8% of respondents provided at least one implausible value, and from 5.57% (Wave 1 of Starting Cohort 4) to 7.01% (Wave 5 of SC3) omitted at least one item. Implausible values, when calculated per item, occurred incidentally, whereas the rates of unspecific missing values per item ranged between 0% and 4.85%, depending on the sample. Item e22683d had visibly higher rates in all samples. The item response distributions of three items were severely skewed and had restricted response scale in selected samples. In another four items, very small percentages of responses were recorded for the lowest category.

In general, the expected three-factor structure appeared in all samples. The three-factor CFA models had a good fit to the data in all but one sample (Wave 3 of SC3). The three-factor EFA model that allowed for cross-loadings had a good fit in the dissimilar sample. The three-factor CFA model for Waves 1 and 5 of SC3 were partial scalar invariant.

The subscales' reliabilities, as measured by Cronbach's α, were marginally acceptable only for the School-Related Factors subscale, but unsatisfactory for the other subscales. The items' explained variances were moderate, indicating that items provided a moderate amount of information about the measured latent response variables.

Since the scale's properties are mixed, data users may face several decisions related its use. These decisions will depend to some extent on the sample they are interested in because the scale's properties differ to some extent between starting cohorts and waves. First, users who intend to use raw scores may decide to omit the Family- and/or Child-Related Factors subscales due to their low reliability and, in the case of Child-Related Factors, due to low item-rest correlations (especially, Wave 5 of SC3 and Wave 1 of SC4). However, users who plan to use latent modelling, which accounts for measurement error, face a decision as to which subscales have acceptable properties in terms of the measurement precision, cross-loadings, or measurement invariance. Moreover, data users need to decide how to deal with cross-loadings in Wave 3 of SC3. Cross-loadings call the validity of the scale into question because it is not entirely clear what cross-loading items measure.

Future changes to the scale may include modifying or replacing severely skewed items, items with low discriminatory power and differentiated properties in various samples. Moreover, making the subscales longer and more homogeneous in content may improve their reliabilities. However, any changes in item content should be preceded by a careful analysis of the theoretical framework. Moreover, it is necessary to gather information on the scale's discriminant and concurrent validity. Such information is not currently available.



Note. SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

Figure 4. Distributions of factor scores derived from the accepted CFA & EFA models.

References

Artelt, C., & Mudiappa, M. (Eds.). (2014). *BiKS - Ergebnisse aus den Längsschnittstudien: Praxisrelevante Befunde aus dem Primar- und Sekundarschulbereich*. Univ. of Bamberg Press. http://nbn-resolving.de/urn:nbn:de:bvb:473-opus4-69584

Bäumer, T., Klieme, E., Kuger, S., Maaz, K., Roßbach, H.-G., Stecher, L., & Struck, O. (2019). Education processes in life-course-specific learning environments. In H.-P. Blossfeld & H.-G. Roßbach (Eds.), *Education as a Lifelong Process—The German National Educational Panel Study (NEPS)* (2nd ed., Vol. 3, pp. 83–99). Springer VS.

Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 25–48). Springer US. https://doi.org/10.1007/978-1-4614-5149-5_2

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*(2), 186–203.

Blossfeld, H.-P., Rossbach, H.-G., & von Maurice, J. (Eds.). (2011). *Education as a Lifelong Process—The German National Educational Panel Study (NEPS). [Special Issue] Zeitschrift fuer Erziehungswissenschaft, 14.*

Codebuch zum Lehrerfragebogen Welle 1 [BiKS-8-14 Grundschule]. (n.d.). Otto-Friedrich-Uuniversität Bamberg, DFG.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Guay, F., Morin, A. J. S., Litalien, D., Valois, P., & Vallerand, R. J. (2015). Application of exploratory structural equation modeling to evaluate the Academic Motivation Scale. *The Journal of Experimental Education*, *83*(1), 51–82. https://doi.org/10.1080/00220973.2013.876231

Hawrot, A. (2021a). *Psychometric properties of the "Quality of instruction" scale in Starting Cohort 3* (NEPS Survey Paper No. 81). Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://doi.org/10.5157/NEPS:SP81:1.0

Hawrot, A. (2021b). *Psychometric properties of the "Teaching quality" scale for teachers in Starting Cohorts 2, 3, and 4* (NEPS Survey Paper No. 82). Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://doi.org/10.5157/NEPS:SP82:1.0

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82. https://doi.org/10.1037/1082-989X.7.1.64

McLarnon, M. J. W., & Carswell, J. J. (2013). The personality differentiation by intelligence hypothesis: A measurement invariance investigation. *Personality and Individual Differences*, *54*(5), 557–561. https://doi.org/10.1016/j.paid.2012.10.029

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(3), 255–273. https://doi.org/10.1080/10705519509540013

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide. Seventh edition*. Muthén & Muthén.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 167–180. https://doi.org/10.1080/10705511.2014.882658

Skopek, J., Pink, S., & Bela, D. (2012). *Data manual. Starting Cohort 3—From lower to upper secondary school. NEPS SC3 1.0.0* (NEPS Research Data Paper). University of Bamberg.

Skopek, J., Pink, S., & Bela, D. (2013). *Starting Cohort 4: 9th Grade (SC4). SUF Version 1.1.0. Data Manual.* (NEPS Research Data Paper). University of Bamberg.

Appendix

Table 1A

Descriptive Statistics of Factor Scores Derived From the Final CFA & EFA Models

Factor	Mean	p50	SD	Var	Skew.	Kurt.	p25	p75	Min	Max
Starting Cohort 3										
Wave 1										
Family	-0.003	-0.098	0.800	0.640	0.009	0.316	-0.558	0.396	-2.434	2.296
Child	-0.018	-0.006	0.674	0.454	-0.165	-0.705	-0.480	0.450	-1.913	1.227
School	-0.014	-0.024	0.773	0.597	0.007	0.283	-0.553	0.599	-2.238	1.356
Wave 3										
Family	0.007	-0.050	0.764	0.584	0.101	0.714	-0.511	0.468	-2.516	2.594
Child	-0.019	-0.018	0.752	0.566	-0.063	-0.350	-0.558	0.522	-2.297	1.621
School	-0.014	-0.098	0.796	0.633	-0.126	-0.355	-0.532	0.613	-2.276	1.702
Wave 5										
Family	-0.002	0.060	0.784	0.614	0.044	0.056	-0.663	0.500	-2.341	2.393
Child	-0.009	-0.091	0.657	0.431	-0.041	-0.384	-0.442	0.484	-1.921	1.562
School	-0.018	0.138	0.765	0.585	-0.286	-0.126	-0.535	0.535	-2.675	1.347
Starting Cohort 4										
Wave 1										
Family	-0.003	-0.046	0.752	0.566	0.058	0.150	-0.572	0.465	-2.559	2.395
Child	-0.024	0.013	0.690	0.476	-0.230	-0.360	-0.558	0.546	-2.514	1.662
School	-0.024	0.040	0.798	0.637	-0.242	-0.403	-0.546	0.636	-2.991	1.314

Note. Family = Family-Related Factors; Child = Child-Related Factors; School = School-Related Factors.