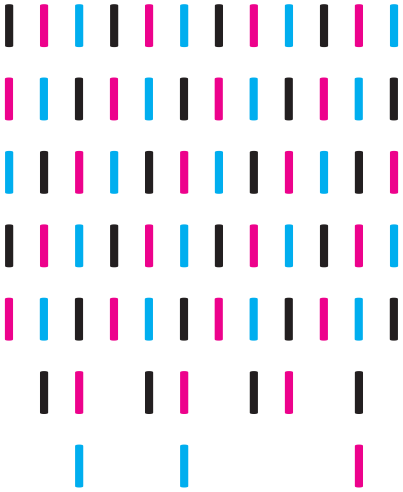


NEPS *SURVEY PAPERS*

Anna Hawrot and Katarina Loos

PROPERTIES OF THE "TEACHER SUPPORT FOR PARENTAL INVOLVEMENT" SCALE FOR TEACHERS IN STARTING COHORT 2, STARTING COHORT 3, and STARTING COHORT 4

NEPS *Survey Paper* No. 88
Bamberg, September 2021

LIfBi
LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

NEPS
National Educational Panel Study

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

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

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

Properties of the "Teacher Support for Parental Involvement" Scale for Teachers in Starting Cohort 2, Starting Cohort 3, and Starting Cohort 4

Anna Hawrot, Katharina Loos
Leibniz Institute for Educational Trajectories, Bamberg, Germany

E-mail address of lead author:

anna.hawrot@lifbi.de

Bibliographic data:

Hawrot, A., & Loos, K. (2021). *Properties of the "Teacher Support for Parental Involvement" Scale for Teachers in Starting Cohort 2, Starting Cohort 3, and Starting Cohort 4* (NEPS Survey Paper No. 88). Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://doi.org/10.5157/NEPS:SP88:1.0

Properties of the "Teacher Support for Parental Involvement" Scale for Teachers in Starting Cohort 2, Starting Cohort 3, and Starting Cohort 4

Abstract

This paper presents information on the source, theoretical background, and psychometric properties of the "Teacher support for parental involvement" scale used Starting Cohorts 2, 3, and 4 to measure if teachers supported parents in being involved in their children's education. We ran an item-level analysis and checked the scale's reliability, internal structure, and measurement invariance across different samples. In general, the items had moderate discriminatory power. Their response distributions were often skewed and the response scale—restricted. The analyses did not confirm the expected unidimensionality of the scale. A two-factor structure emerged, but this required excluding two items. Internal consistencies of the identified subscales ranged from low (approx. .5 or less) for the "Attitude" subscale to acceptable for "Communication" (approx. .7 or more). The scale was scalarly invariant across the waves of Starting Cohort 3, as well as across Starting Cohort 3 and Starting Cohort 2. Overall, the scale has mixed psychometric properties and needs refinements.

Keywords

parental involvement, psychometric properties, teachers

Acknowledgments

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 2 (doi: 10.5157/NEPS:SC2:7.0.0), 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.

We 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 (Blossfeld et al., 2011) aims at tracking the development of various competencies, describing their patterns, and better understanding how they unfold across the lifespan. 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 their developmental and educational stage.

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 "Teacher support for parental involvement" scale used in Starting Cohorts 2, 3, and 4 to assess if teachers supported parents in being involved in their children's education. Its goal 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. Description of the Scale

The "Teacher support for parental involvement" scale measures to what extent the school and teachers in their everyday pedagogical work support parents in being involved in their children's education. In other words, it focuses on school and teacher behavior that invites parents to be engaged in their child's educational endeavors and supportive of school goals and school learning (Gerecht et al., 2007; Steinert et al., 2003).

The scale is assumed to be unidimensional and consists of 8 items. Respondents are asked to rate to what extent each item reflects their personal opinion about cooperation with parents. The response options are as follows: 1 = *does not apply at all*, 2 = *does rather not apply*, 3 = *does rather apply*, and 4 = *applies completely*. The scale was developed in the project 'Schulentwicklung, Qualitätssicherung und Lehrerarbeit' (SEL; Steinert et al., 2003). However, items e22684c, e22684f, and e22684g were slightly modified for the purpose of the National Educational Panel Study.

Table 1 contains the item wording and the corresponding variable names used in the Scientific Use Files. The original German-language version of the scale 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 – Teacher attitudes".

Table 1

Items of the "Teacher Support for Parental Involvement" Scale

Variable name	Firstly, please tell us your personal opinion with regard to teaching and learning
e22684a	a) I like working with parents.
e22684b	b) I consider parents as partners in educating and raising their children.
e22684c	c) I keep my students' parents updated on a regular basis about what's going on at school.
e22684d	d) I follow up on parent complaints/concerns.
e22684e	e) I tell parents about the strengths and weaknesses of their children.
e22684f	f) I tell parents about the learning progress of their children on a regular basis.
e22684g	g) Parents can make an appointment to see me at the school to discuss their children's issues at school.
e22684h	h) Parents can talk to me about their children's issues at school even outside of school hours.

3. Method

3.1 Data and Sample

We used data gathered during the National Educational Panel Study (NEPS, Blossfeld et al., 2011) from Starting Cohort 2 (SC2), Starting Cohort 3 (SC3) and Starting Cohort 4 (SC4). In subsequent waves of each cohort, selected teachers who thought sampled students were surveyed as context persons. The scale was administered as a part of a larger questionnaire using the standard testing procedure for a wave. Information on the procedure is available in the data manuals (Skopek et al., 2012a, 2012b, 2013) and interviewer manuals[1].

Table 2 contains information about the waves and starting cohorts in which teachers filled in the scale. It is supplemented by information which grade students participating in NEPS attended at a given measurement occasion. Please note that teachers employed in special schools were excluded from the sample.

[1] https://www.neps-data.de/Data-Center/Data-and-Documentation/Starting-Cohort-Kindergarten/Documentation; 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 2, 3, and 4

Wave	3	4	5	6	7
Starting Cohort 2			G3 (2014/15)		
Starting Cohort 3		G8 (2013/14)			G10 (2015/16)
Starting Cohort 4	G10 (2011/12)				

Note. G = grade. School years are provided in parentheses.

In SC2 and SC4, teachers filled in the scale once only. However, in SC3 the scale was administered twice to all sampled teachers. As a consequence, the same teacher could fill in the scale twice: in Wave 4 and in Wave 7. 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 administration only, teacher samples in Waves 4 and 7 of SC3 did not overlap. We used variable *ex80211*, which contains information about the questionnaire administered to each teacher (first-time or panel interviewee questionnaire), to identify and exclude all repeated administrations.

Table 3 presents the sample sizes in all of the 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 Administration Time-Points

Wave	3	4	5	6	7
Starting Cohort 2			642		
Starting Cohort 3		765			278
Starting Cohort 4	1073				

Note. Samples include first-time administrations 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, for instance, range restrictions. Next, we checked items' discriminatory power.

The second step involved analyzing the internal structure of the scale and was divided into three sub-steps; first, confirmatory factor analyses (CFA) were performed for each sample

separately; then, since CFA models did not provide an adequate fit to the data, exploratory factor analyses (EFA) were performed.

The third step involved verifying measurement invariance of the most optimal measurement model across (a) the two waves of SC3, (b) SC2 and SC3, and (c) SC2 and SC4. Samples in subsequent waves of each starting cohort included the first administration of the scale, therefore they were treated as independent groups. However, SC3 and SC4 samples were drawn from the same schools, thus they could overlap. Although each teacher surveyed in both cohorts should have one and the same identification number, we could not assure that because of the above mentioned inconsistencies in gender and birth date of teachers having the same identification number assigned. As a consequence of the difficulties in identifying dependent data, we decided not to test invariance across SC3 and SC4.

We tested configural, metric, and scalar invariance within the exploratory structural equation modeling (ESEM) framework using Millsap & You-Tein's (2004) model specifications and delta parameterization. The models were compared using the DIFFTEST procedure (Muthén & Muthén, 1998-2015). We did not use ΔAFIs for comparison purposes because they are not recommended with the WLSMV estimator (Sass et al., 2014). Please note that it is not allowed to free loadings in the ESEM framework. As a result, we decided to test factor loadings and thresholds not in tandem but separately. This approach has been 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 of the analyses of internal structure were performed with Mplus 7.4 (Muthén & Muthén, 1998-2015) using 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). Initial EFA models used oblique Geomin rotation, whereas the later ones—oblique target rotation (e.g., Browne, 2001). We used target rotation after having decided how many factors should be extracted. 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: 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 did not exceeded 1, and under 0.4% of respondents participating in a wave provided at least one implausible response.

The number of unspecific missing values per person was higher than the number of implausible values. The percentage of respondents with one unspecific missing value or more varied between 2.16% in Wave 7 of SC3 and 3.89% in Wave 5 of SC2. The respondents most often omitted one item, and the rate of single-item omissions varied between 1.44% (Wave 7 of SC3) and 3.12% (Wave 5 of SC2).

The total missing values per person and the number 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	**SC2: W5**		**SC3: W4**		**SC3: W7**		**SC4: W3**	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
Implausibe value								
0	641	99.84	762	99.61	278	100	1071	99.81
1	1	0.16	3	0.39	0	0	2	0.19
2	0	0	0	0	0	0	0	0
Total	642	100	765	100	278	100	1073	100
Unspecific missing								
0	617	96.11	745	97.39	272	97.84	1042	97.11
1	20	3.12	17	2.22	4	1.44	26	2.42
2	2	0.31	1	0.13	1	0.36	1	0.09
3	2	0.31	1	0.13	1	0.36	0	0
4	0	0	0	0	0	0	2	0.19
5	1	0.16	1	0.13	0	0.00	2	0.19

Number of missing values per person	SC2: W5		SC3: W4		SC3: W7		SC4: W3	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
6	0	0	0	0	0	0	0	0
Total	642	100	765	100	278	100	1073	100
Total missing								
0	616	95.95	742	96.99	272	97.84	1040	96.92
1	21	3.27	20	2.61	4	1.44	28	2.61
2	2	0.31	1	0.13	1	0.36	1	0.09
3	2	0.31	1	0.13	1	0.36	0	0
4	0	0	0	0	0	0	2	0.19
5	1	0.16	1	0.13	0	0	2	0.19
6	0	0	0	0	0	0	0	0
Total	642	100	765	100	278	100	1073	100

Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

4.1.2 Missing responses per item

Table 5 contains information about implausible, unspecific, and total missing values per item in all samples. Implausible values occurred incidentally. Unspecific missing value rates per item were also low and ranged from 0% to 1.87% depending on the item and sample. Items e22684a and e22684f showed slightly higher rates in comparison to the other items, but due to generally low rates the difference was very small.

The number of total missing values per item and the number 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 total missing values are not described.

Table 5

Rates of Missing Values per Item

Variable name	SC2: W5		SC3: W4		SC3: W7		SC4: W3	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Implausibe value								
e22684a	0	0	0	0	0	0	0	0
e22684b	1	0.16	0	0	0	0	0	0
e22684c	0	0	0	0	0	0	2	0.19
e22684d	0	0	1	0.13	0	0	0	0
e22684e	0	0	2	0.26	0	0	0	0
e22684f	0	0	0	0	0	0	0	0
e22684g	0	0	0	0	0	0	0	0
e22684h	0	0	0	0	0	0	0	0
Unspecific missing								
e22684a	12	1.87	9	1.18	1	0.36	11	1.03
e22684b	3	0.47	3	0.39	1	0.36	3	0.28
e22684c	1	0.16	5	0.65	3	1.08	9	0.84
e22684d	3	0.47	3	0.39	1	0.36	4	0.37
e22684e	1	0.16	2	0.26	1	0.36	5	0.47
e22684f	7	1.09	4	0.52	2	0.72	8	0.75
e22684g	4	0.62	0	0	0	0	3	0.28
e22684h	4	0.62	1	0.13	0	0.00	3	0.28

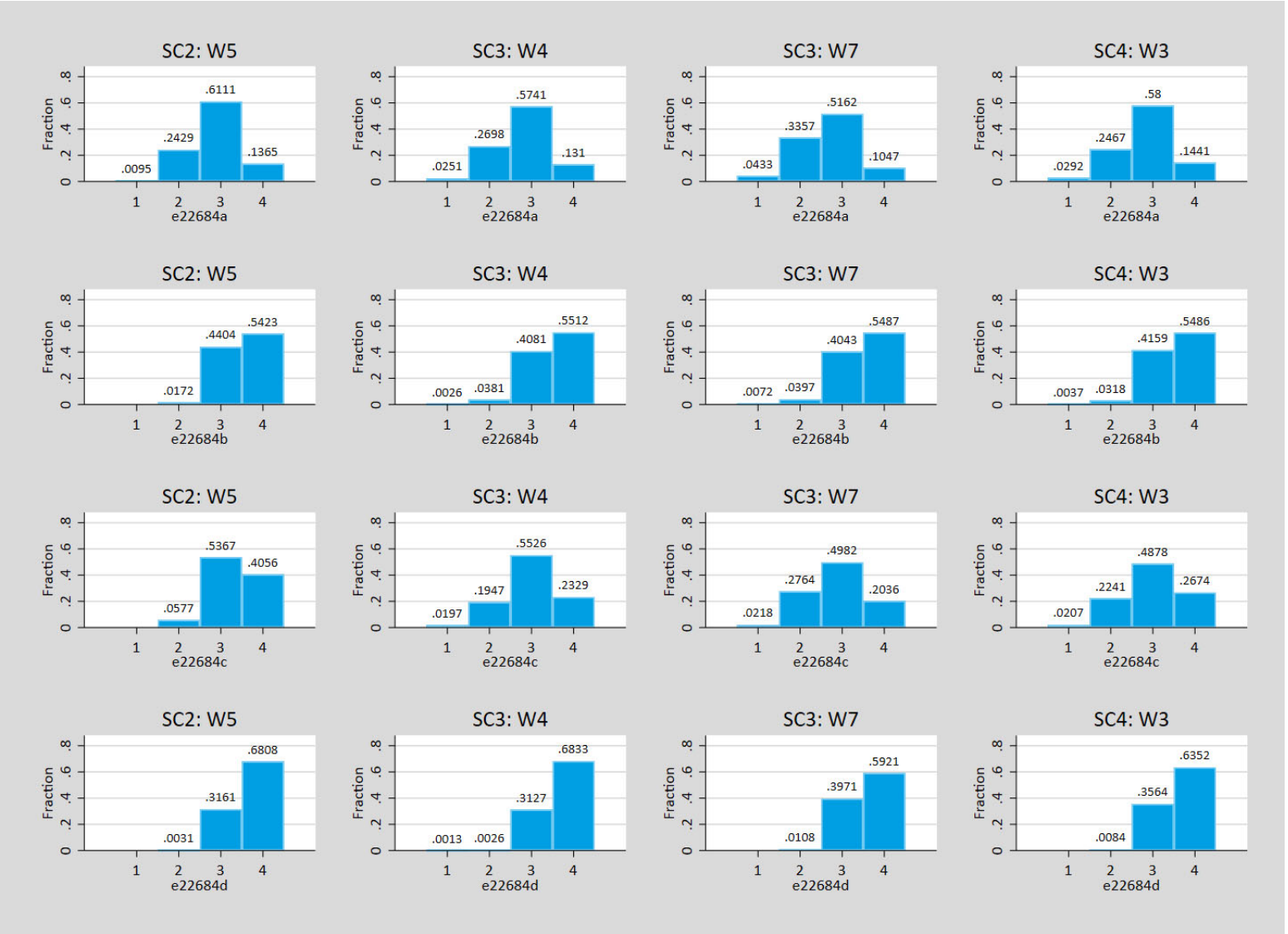
Variable name	SC2: W5		SC3: W4		SC3: W7		SC4: W3	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Total missing								
e22684a	12	1.87	9	1.18	1	0.36	11	1.03
e22684b	4	0.62	3	0.39	1	0.36	3	0.28
e22684c	1	0.16	5	0.65	3	1.08	11	1.03
e22684d	3	0.47	4	0.52	1	0.36	4	0.37
e22684e	1	0.16	4	0.52	1	0.36	5	0.47
e22684f	7	1.09	4	0.52	2	0.72	8	0.75
e22684g	4	0.62	0	0.00	0	0.00	3	0.28
e22684h	4	0.62	1	0.13	0	0.00	3	0.28

Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

In summary, the implausible value rates per item were negligible. In conjunction with the negligible rates per person, this result suggests that respondents did not experience major difficulties in using the scale's response format. The unspecific missing value rates per item and per person were also highly satisfactory.

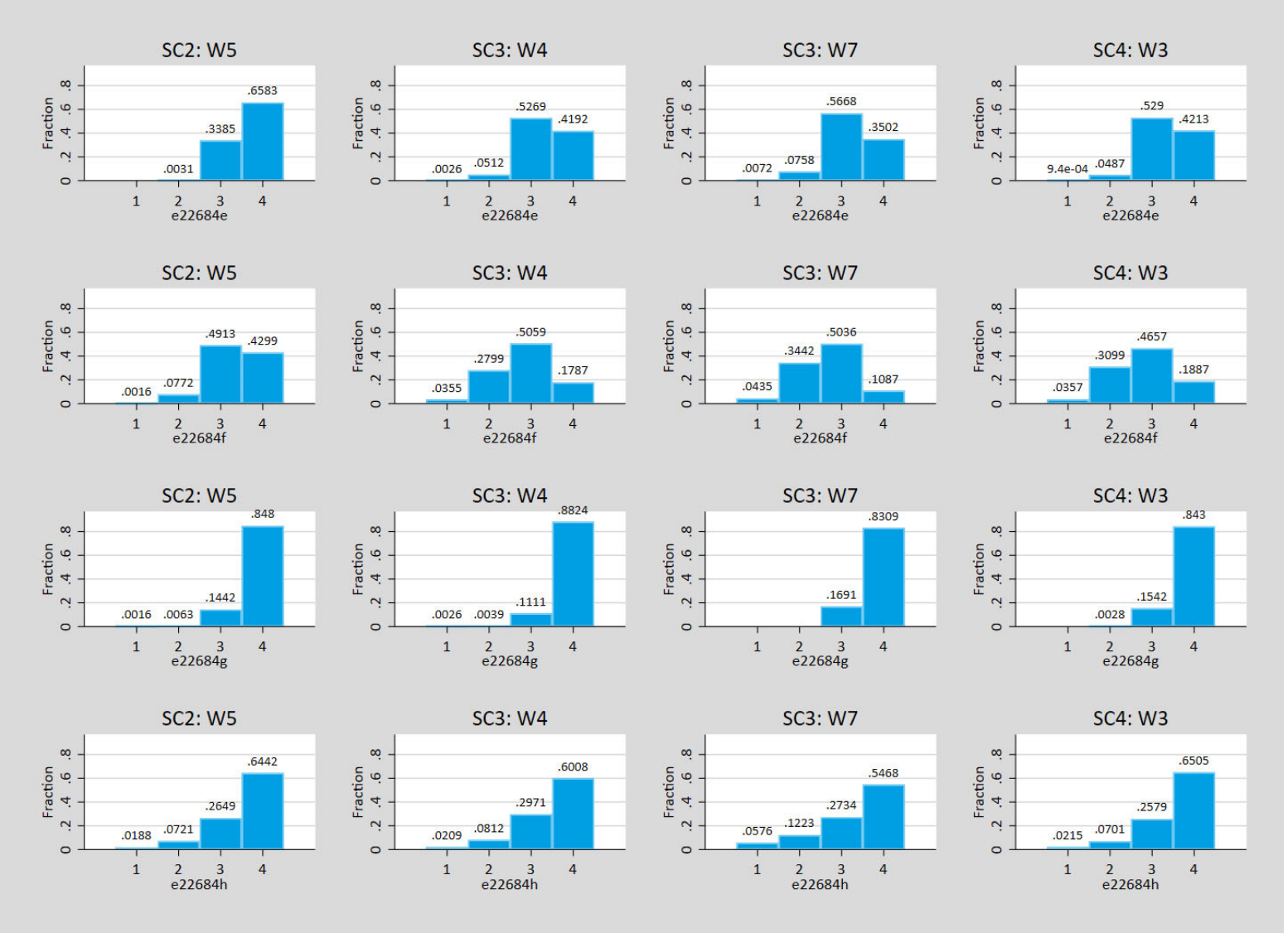
4.2 Item Distributions

Figures 1 and 2 present the item response distributions in each sample. Their analysis showed three unfavorable properties of the items. First, the distributions were skewed or severely skewed. Second, ceiling effects appeared (see e.g., e22684b, e22684d, e22684g). Third, restrictions of the response scale occurred. No responses were recorded for the category *does not apply at all* (1) for 5 items in 8 out of 32 item-sample combinations. Moreover, in Wave 7 of SC3 no responses were recorded for two categories (*does not apply at all*, *does not really apply*) for item e22684g. Even if scale restrictions did not occur, the lowest category was rarely chosen (e.g., items e22684a, e22684f, e22684h).



Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave; 1 = does not apply at all, 2 = does rather not apply, 3 = does rather apply, and 4 = applies completely.

Figure 1. Item response distributions: items e22684a - e22684d.



Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave; 1 = does not apply at all, 2 = does rather not apply, 3 = does rather apply, and 4 = applies completely.

Figure 2. Item response distributions: items e22684e - e22684h.

4.3 Discriminatory Power

To assess the discriminatory power of the items, we calculated item-rest correlations (Spearman's rho). In general, their values, which ranged between .24 and .57, were satisfactory. However, one item (e22684h) had relative lower discriminatory power in all samples. Moreover, only one item (e22684e) showed high discriminatory power (>= .5) in all samples. The discriminatory power for the remaining items oscillated around moderate to high values. The results are presented in Table 6.

Table 6

Item-Rest Correlations (Spearman's rho)

Item	SC2: Wave 5	SC3: Wave 4	SC3: Wave 7	SC4: Wave 3
e22684a	.331	.439	.356	.414
e22684b	.515	.420	.407	.450
e22684c	.520	.502	.420	.566
e22684d	.559	.413	.490	.491
e22684e	.556	.551	.519	.583
e22684f	.430	.507	.383	.524
e22684g	.449	.298	.350	.369
e22684h	.300	.240	.297	.312

Note. SC = Starting Cohort.

4.4 Internal Structure

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

In the first step, we ran confirmatory factor analysis for each sample separately to test whether the expected one-factor structure held in independent samples. The models did not include any cross-loadings. All of the models showed a poor fit; detailed information is presented in Table 7.

Table 7

Fit of the One-Factor CFA Models

Wave	*n*	N_{par}	χ^2	*df*	*p*	RMSEA	CFI	TLI
Starting Cohort 2								
5	639	28	224.19	20	< .001	.126	.933	.907
Starting Cohort 3[a]								
4	763	29	360.83	20	< .001	.149	.860	.803
7	277	29	109.34	20	< .001	.127	.862	.806
Starting Cohort 4								
3	1069	30	489.72	20	< .001	.148	.884	.838

Note. N_{par} = number of free parameters. Number of free parameters differs between starting cohorts because of the differences in response scale restrictions.

[a] Response categories *does not apply at all* (1) and *does rather not apply* (2) were merged for item e22684d. Response categories *does not apply at all* (1), *does rather not apply* (2), and *does rather apply* (3) were merged for item e22684g. This assured the same number of categories for these items in both waves of SC3.

Second, since the CFA models had an unsatisfactory fit, we explored the scale's dimensionality by running exploratory factor analysis (EFA)[2] in each sample separately. We tested models having up to three factors. Model fit indices, scree plots, Kaiser criterion, and interpretability of the results served as a basis for determining the number of factors. The selection criteria suggested a two-factor solution in all samples, however, only one two-factor model showed an acceptable fit to the data. Based on a detailed inspection of the model for SC2, W5, we excluded items e22684d and e22684g, and retested the two-factor solution[3] for that sample. The model showed a good fit to the data, and therefore we rerun it on the remaining samples[4]. Details are presented in Table 8, the scree plots are included in the Appendix (Figure 1A).

The inspection of factor loadings in the final EFA models (see Table 9) revealed that the items usually loaded on either "Attitude" or "Communication" factor, although items e22684c and e22684e cross-loaded in all samples. Cross-loadings were particularly high (between 0.3 and 0.37) in Wave 5 of SC2 and Wave 7 of SC3. Despite this, the internal structure was similar in all waves and starting cohorts.

[2] Geomin rotation.

[3] Target rotation.

[4] Target rotation.

Table 8

Fit of the Final Two-Factor EFA Models

Wave	*n*	N_{par}	X^2	*df*	*p*	RMSEA	CFI	TLI
Starting Cohort 2								
5	639		12.095	4	.017	.056	.995	.980
Starting Cohort 3								
4	736	11	7.069	4	.132	.032	.999	.994
7	277	11	2.150	4	.708	.000	1.00	1.02
Starting Cohort 4								
3	1069	11	4.868	4	.301	.014	1	.999

Note. N_{par} = number of free parameters.

Table 9

Factor Loadings and Factor Correlations of Final Two-Factor EFA Models

Item	SC2: Wave 5		SC3: Wave 4		SC3: Wave 7		SC4: Wave 3	
	F1	F2	F1	F2	F1	F2	F1	F2
e22684a	**0.643***	-0.111	**0.765***	0.025	**0.489***	0.100	**0.721***	-0.001
e22684b	**0.816***	0.019	**0.722***	-0.010	**0.794***	-0.009	**0.710***	0.011
e22684c	0.374*	**0.441***	0.218*	**0.584***	0.324*	**0.431***	0.203*	**0.630***
e22684e	0.306*	**0.674***	0.183*	**0.651***	0.360*	**0.529***	0.156*	**0.727***
e22684f	-0.005	**0.943***	-0.009*	**0.955***	-0.004	**0.960***	-0.022	**0.911***
e22684h	**0.457***	-0.104	**0.355***	-0.037	**0.258***	0.065	**0.498***	-0.091
Corr.	.402*		.473*		.307*		.507*	

Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; Corr. = factor correlation; F1 = factor 1; F2 = factor 2. Loadings of items assigned to a given factor are in bold type.
* $p < .05$.

The salient factor loadings varied, ranging from 0.258 (e22684h, Wave 7 of SC3) to 0.960 (e22684f, Wave 7 of SC3). Only two salient loadings had values below 0.4, whereas 17 (70.8%) equaled 0.5 or more, and 15 (62.5%) 0.6 or more. The loadings of items e22684b and e22684f

equaled 0.7 or higher in all samples. The factor correlations were moderate. Factor loadings and factor correlations in the final two-factor solutions are available in Table 9.

In summary, a two-factor structure emerged in all of the samples, however, several items cross-loaded on the other dimension. The magnitude of salient factor loadings, although differentiated, was acceptable. The magnitude of cross-loadings was marginally acceptable.

4.4.1 Measurement invariance

We checked the scale's measurement invariance across (a) the two waves of the SC3, (b) SC3 and SC2, and (c) SC2 and SC4. Table 10 contains a summary of the results. The scale was invariant across the two Waves of the SC3, and across SC3 and SC2. However, is showed configural invariance only and across SC2 and SC4. Please note that the ESEM framework does not allow to free factor loadings, thus establishing partial invariance was impossible.

Table 10

Results of Measurement Invariance Testing

model	N_{par}	RMSEA	CFI	TLI	χ^2	*Df*	*p*	Comp.	$\Delta\chi^2$ *(df), p*[d]
Starting Cohort 3[a]									
(1)	58	.013	1.000	0.999	8.66	8	.372		
(2)	50	.000	1.000	1.003	10.25	16	.726	2 vs 1	4.62 (8), *p* = .797
(3)	40	.000	1.000	1.000	25.24	26	.506	3 vs 2	13.31 (10), *p* = .207
Starting Cohort 3 and Starting Cohort 2[b]									
(1)	78	.035	0.998	0.992	20.46	12	.059		
(2)	62	.033	0.996	0.993	44.83	38	.024	2 vs 1	25.87 (16), *p* = .056
(3)	48	.030	0.995	0.994	63.84	42	.017	3 vs 2	19.72 (14), p = .14
Starting Cohort 4 and Starting Cohort 2[c]									
(1)	52	.038	0.998	0.991	17.93	8	.022		
(2)	44	.043	0.994	0.989	40.89	16	< .001	2 vs 1	22.45 (8), *p* = .004

Note. (1) = configural; (2) = metric; (3) = scalar; N_{par} = number of parameters; Comp. = compared models.
[a] N_{Wave4}: 763, N_{Wave7}: 277.
[b] SC3: N_{Wave4}: 763, N_{Wave7}: 277; SC2: N_{Wave5}: 639. Two adjacent response categories of items e22684b, e22684c, e22684e, i.e., *does not apply at all* (1) and *does rather not apply* (2), were merged to assure the same number of response categories across the tested waves.
[c] SC4: N_{Wave3}: 1069; SC2: N_{Wave5}: 639. Two adjacent response categories of items e22684b, e22684c, e22684e, i.e., *does not apply at all* (1) and *does rather not apply* (2), were merged to assure the same number of response categories across the tested waves.
[d] DIFFTEST procedure was used.

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 final 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 11 presents the items' explained variance in all samples[5]. They differed between the items and samples, ranging from .081 to .919. A total of 18 out of 24 item-sample combinations had values of .5 or higher. This indicates that a moderate amount of the items' variance was accounted for in the models. Item e22684h showed particularly low values.

Table 11

Items' Explained Variances (R^2) in the Final Two-Factor EFA Models

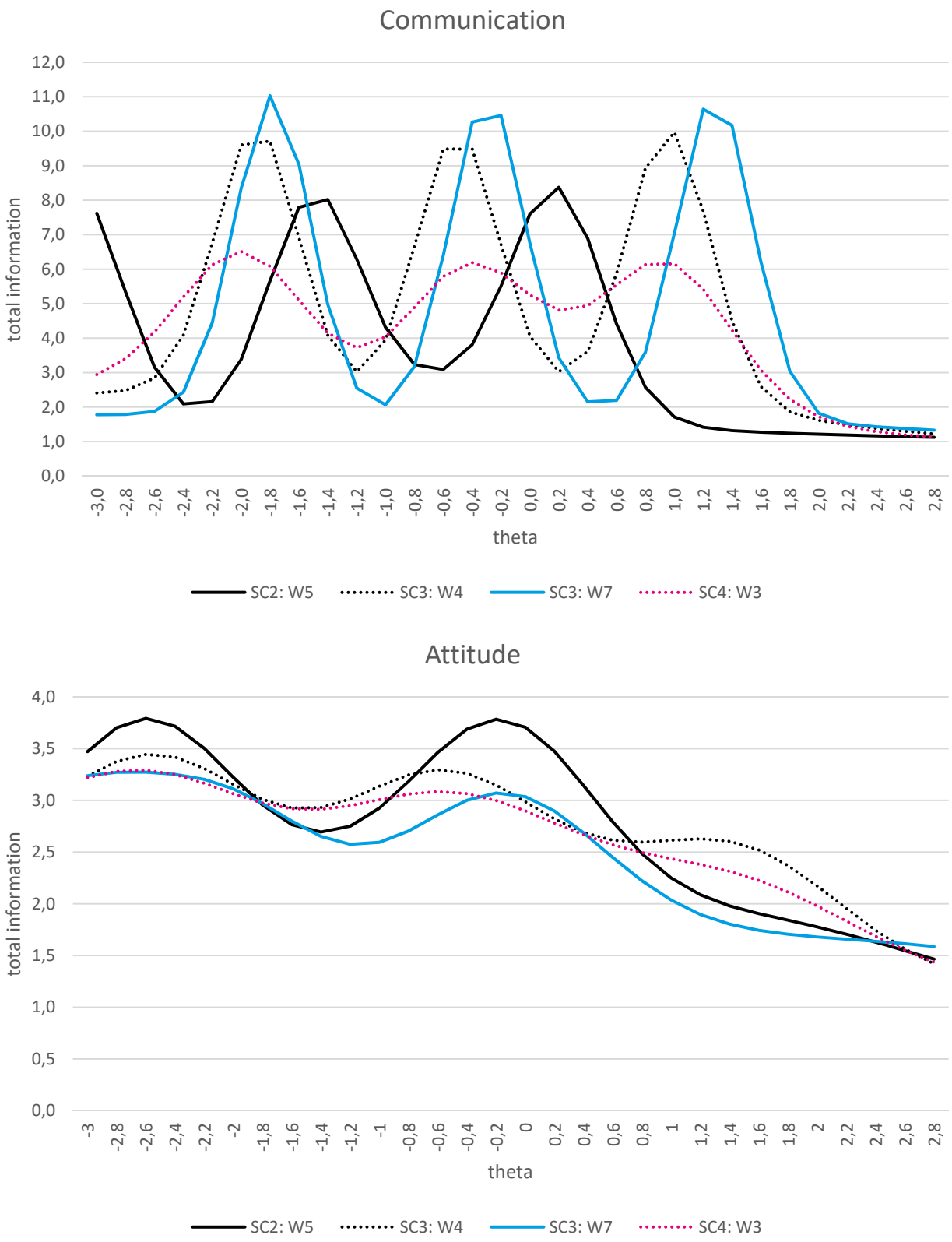
Item	SC2	SC3		SC4
	Wave 5	Wave 4	Wave 7	Wave 3
e22684a	.369	.604	.279	.520
e22684b	.679	.515	.626	.513
e22684c	.467	.509	.377	.568
e22684e	.714	.569	.527	.669
e22684f	.885	.904	.919	.811
e22684h	.182	.115	.081	.211

Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4.

Figure 3 presents the total information curves of both factors in all of the samples. The measurement precision of the Communication subscale differed between various trait levels. The observed pattern is probably a result of the fact that the subscale consists of only three items, but the items differ in difficulty. The Attitude subscale's precision showed a decreasing trend with an increasing trait level.

Table 12 contains information on the internal consistency of the scale. The reliability of the Attitude subscale was unsatisfactory: it ranged between .38 and .54. However, it ranged between .69 and .78 for the Communication subscale and therefore was acceptable.

[5] 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. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

Figure 3. Total information curves of the EFA factors.

Table 12

Cronbach's α Coefficients

Subscale	SC2	SC3		SC4
	Wave 3	Wave 4	Wave 7	Wave 3
Attitude	.493	.511	.377	.536
Communication	.715	.760	.688	.777

Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4.

Since Cronbach's α depends to some extent on the number of items, its low values are common in short scales. Moreover, high reliability of a short scale may suggest that the scale taps only selected aspects of a given construct and although the items evoke consistent responses, the scale may be of limited validity.

The analysis of the item content suggests that heterogeneity of the items might indeed affect the subscales' reliabilities. The Attitude subscale covers the affective (e22684a), cognitive (e22684b), and behavioral (e22684h) component of an attitude, whereas the Communication subscale focuses on a single aspect of teacher-parent communication (topics). As a consequence, the latter subscale is much more internally consistent.

However, the internal consistency as low as .4 or .5 suggests that the Attitude subscale needs a careful revision and in present form should be used with caution. It may be necessary to increase the number of items to better cover each component or even create a separate subscale for each component, depending on the intended use of the scale.

4.6 Factor Scores Distributions

Figure 4 presents the distributions of the factor scores derived from the final EFA models. Various deviations from normality were present, including non-symmetry (e.g., the Attitude factor in Wave 5 of SC2), ceiling effects (e.g., the Communication factor in Wave 5 of SC2), or multi-modality (e.g., the Attitude factor in Wave 7 of SC3). Descriptive statistics are available in the Appendix (Table 1A).

5. Summary

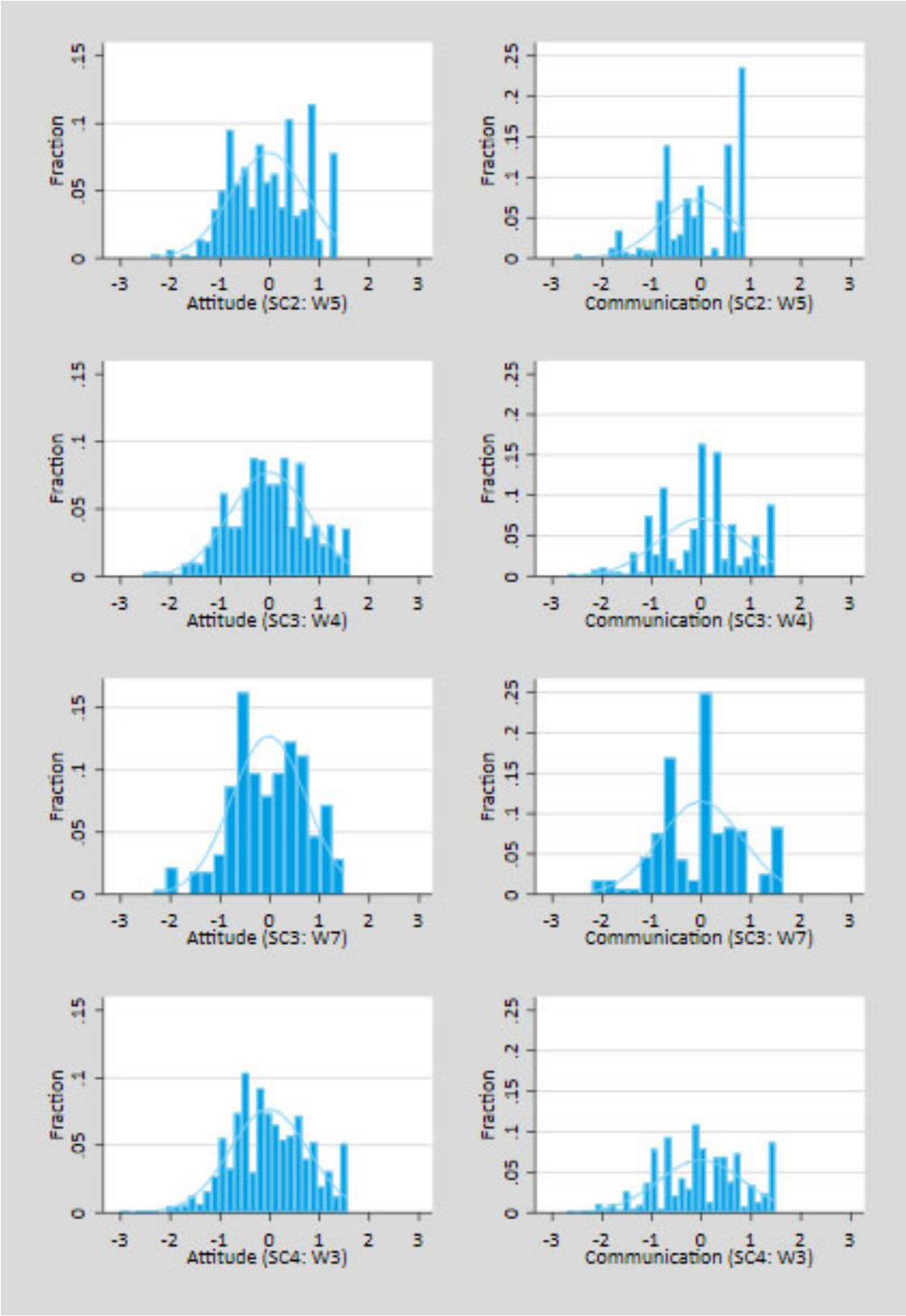
This paper documents the "Teacher support for parental involvement" scale used in the National Educational Panel Study to measure if teachers supported parents in being involved in their children's schooling. Besides providing information about the scale's source and theoretical background, the paper reports basic information about its psychometric properties.

The scale was administered to teachers teaching the sampled students in Starting Cohort 2, Wave 5, Starting Cohort 3, Waves 4 and 7, and Starting Cohort 4, Wave, 3. The analyses included 2,758 teachers in total. They revealed that the missing values rates per person and per item were highly satisfactory. In each sample under 0.5% of respondents provided at least one implausible response, and under 3.89% omitted at least one item. The unspecific missing value rates per

item did not exceed 1.87%. However, the item response distributions were highly skewed and response scale restrictions occurred for in the case of 5 items. In the most severe case, responses for only two response categories were recorded (e22684g in SC3, W7).

The one-factor CFA model, representing the expected internal structure, did not fit to the data. The exploratory factor analyses revealed a two-dimensional oblique structure of the scale, but this required excluding two items. The final measurement model showed scalar invariance across the two Waves of SC3, as well as SC3 and SC2. However, only configural invariance held across SC2 and SC4. As a consequence, it is not recommended to directly compare the two starting cohorts.

The subscales' reliabilities as measured by Cronbach's α were acceptable for the Communication subscale (about .7 or higher), but unsatisfactory for the Attitude subscale (about .5 or less). The items' explained variances were differentiated; they ranged between .081 and .919 depending on the item and sample. In general, the scale requires major refinements. Potential modifications include replacing the items with skewed distributions, low factor loadings, and cross-loadings as well as clarifying the scale's theoretical background (and as a consequence—its expected dimensionality).



Note. SC2 = Starting Cohort 2; SC3 = Starting Cohort 3; SC4 = Starting Cohort 4; W = Wave.

Figure 4. Distributions of factor scores derived from the final two-factor EFA models.

References

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.

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.*

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05

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

Gerecht, M., Steinert, B., Klieme, E., & Döbrich, P. (2007). *Skalen zur Schulqualität. Dokumentation der Erhebungsinstrumente. Pädagogische Entwicklungsbilanzen mit Schulen (PEB)* (Vol. 17). GFPF.

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

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*(3), 479–515. https://doi.org/10.1207/S15327906MBR3903_4

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

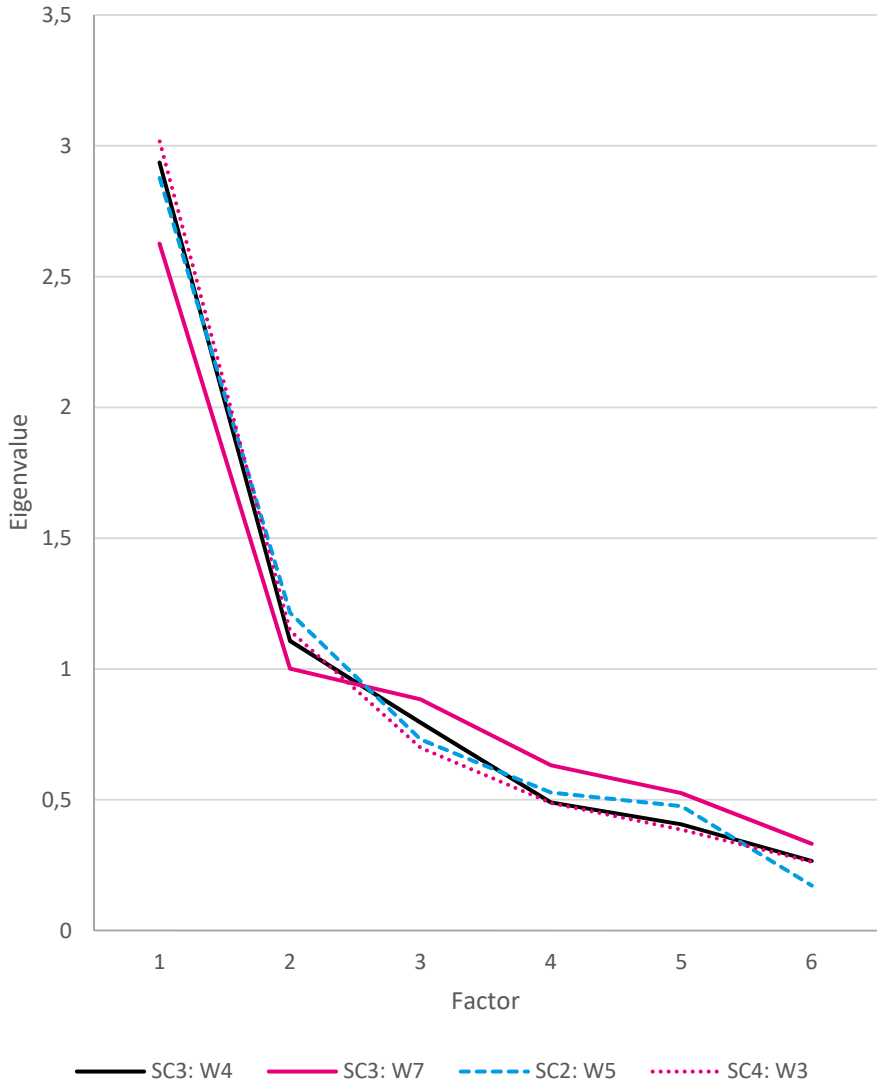
Skopek, J., Pink, S., & Bela, D. (2012a). *Data manual. Starting Cohort 2—From kindergarten to elementary school. NEPS SC2 1.0.0* (NEPS Research Data Paper). University of Bamberg.

Skopek, J., Pink, S., & Bela, D. (2012b). *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.

Steinert, B., Gerecht, M., Klieme, E., & Döbrich, P. (2003). *Skalen zur Schulqualität: Dokumentation der Erhebungsinstrumente. ArbeitsPlatzUntersuchung (APU), Pädagogische EntwicklungsBilanzen (PEB)* (Vol. 10). GFPF.

Appendix



Note. W = Wave.

Figure 1A. Eigenvalues of the extracted EFA factors in Starting Cohort 2 (SC2), 3 (SC3), and 4 (SC4).

Table 1A

Descriptive Statistics of Factor Scores Derived From the Final Two-Factor EFA Models

Factor	Mean	p50	SD	Var.	Skew.	Kurt.	p25	p75	Min	Max
Starting Cohort 2										
Wave 5										
Attitude	-0.026	-0.039	0.764	0.583	-0.051	-0.644	-0.654	0.552	-2.380	1.363
Activites	-0.067	-0.073	0.760	0.577	-0.492	-0.544	-0.734	0.710	-2.551	0.881
Starting Cohort 3										
Wave 4										
Attitude	-0.016	-0.006	0.796	0.634	-0.122	-0.158	-0.492	0.570	-2.551	1.606
Activites	-0.018	-0.032	0.857	0.735	-0.212	-0.384	-0.747	0.579	-2.678	1.465
Wave 7										
Attitude	-0.022	0.017	0.751	0.564	-0.257	-0.194	-0.574	0.536	-2.337	1.491
Activites	-0.012	0.124	0.834	0.696	-0.060	-0.105	-0.625	0.517	-2.199	1.642
Starting Cohort 4										
Wave 3										
Attitude	-0.017	-0.024	0.792	0.627	-0.115	-0.077	-0.532	0.566	-3.004	1.559
Activites	-0.013	-0.027	0.858	0.737	-0.170	-0.477	-0.683	0.644	-2.697	1.487