Transitionen von der Erstausbildung ins Erwerbsleben
Transitions de l'Ecole à l'Emploi
Transitions from Education to Employment

UNIVERSITÄT
BERN

Scaling methodology and scale reporting in the TREE2 panel survey

Documentation of scales implemented in the baseline survey (2016)

(Update 2023)

Stefan Sacchi
Dominique Krebs-Oesch

Imprint

Published by TREE (Transitions from Education to Employment).
University of Bern
Fabrikstr. 8
3012 Bern/Switzerland

www.tree.unibe.ch

tree.soz@unibe.ch

Suggested citation

Sacchi, Stefan, Krebs-Oesch, Dominique (2023). Scaling methodology and scale reporting in the TREE2 panel survey. Documentation of scales implemented in the baseline survey (2016). Update 2023. Bern: TREE. doi 10.48350/179867

Abstract

This documentation refers to the database of the 2nd TREE cohort's (TREE2) as published in the 2023 data release (TREE, 2023). It outlines the statistical models and estimation methods employed for scale construction and the calculation of student scores based on questionnaire items. Furthermore, we discuss the various metrics and indicators of relevant scale properties compiled in the technical appendix for all scales implemented in the TREE2 baseline survey.

The focus of the scale reporting is on the internal consistency of the scales and on the comparability of the measurements across survey languages, survey modes and survey settings involved. With very few exceptions, the results indicate at least sufficient or high internal consistency and measurement invariance of the scales used.

A complementary documentation covering the scales employed in later panel waves can be found in the 2023 TREE2 data release (Sacchi & Krebs-Oesch, 2023). With the exception of a few additional metrics of longitudinal measurement invariance over panels waves (ibid., sections 3.6, 4.2), it basically relies on the methods presented below.

Table of Contents

Figures

Tables

Some practical guidelines for using the scales

For each scale administered in the TREE2 baseline survey, the technical appendix of this documentation provides a selection of relevant scale metrics and quality measures. Section 4 of the introductory text describes the type and calculation of the reported measures and gives some clues as to their interpretation. We thus intend to support data users in assessing measurement properties of the scales in question. Note that for some of the scales administered in the baseline survey, one or more repeated measurements from later panel waves are available, which are documented in Sacchi & Krebs-Oesch (2023).

The reported scale-specific measures focus primarily on reliability (in the sense of internal consistency) and measurement invariance across survey settings, modes and languages. What we do not address in this documentation is scale validity, as TREE mostly uses commonly accepted, well-established scales and validity is therefore not likely to be a major problem. In addition, the database offers researchers many opportunities to conduct external validations tailored to their specific analytical needs.

In some cases, several scales in the TREE2 scientific use file partly draw on one and the same items. The scales in question should therefore not be used simultaneously within the same multivariate model. This concerns some scales for which several versions exist (cf. section 2: scales surrounded by dotted lines in Table 3) as well as other scales composed of main and subdimensions (cf. section 2, Table 4).

Regarding the use of student scores in the context of multivariate models, we refer the reader to the remarks on this issue in section 3.2.2. Some scores represent item composites rather than scale scores (cf. Table 5), which may, however, be used similarly. The variable names (short names without wave-specific prefix) and labels of all items, student scores and composite variables in the technical appendix correspond with those in the TREE2 data release (TREE, 2023).

When estimating the confirmatory factor models and calculating the student scores, we imputed all missing item information, provided that at least one item of a given scale had a valid rating (see section 3.1.1b for details).

Introduction

This paper documents the questionnaire-based scales and item-based composites that have been collected on the occasion of the baseline survey administered to the second TREE cohort (TREE2) in 2016. First, the paper focuses on the methods and the estimation procedures that we have adopted for the calculation of the scale values published in the scientific use data files. Second, we describe the calculation of the scale-specific key figures and quality parameters (see appended tables) and provide some useful information for their interpretation.

The TREE2 baseline survey is composed of two surveys carried out at a short interval in spring/summer 2016. The first survey is a large-scale national assessment of mathematics skills administered to students who had reached the end of compulsory school (Assessment of the Attainment of Educational Standards, henceforth AES).[1] Beyond the assessment itself, the AES survey programme included a comprehensive student background questionnaire that collected a wide range of student background characteristics presumed to influence maths skills development and/or educational and labour-market pathways in the further (post-compulsory) life course. The second survey, which we refer to as extension survey, was conducted shortly after the first one. Its main purpose was to complete some student background characteristics that had not been collected among all respondents of the first survey. In doing so, TREE was able to substantially extend the size of the TREE2 starting cohort (see section 1 for details).

All parts of the AES student questionnaire include numerous item-based measures designed to capture latent (i.e., not directly observable) respondent, family or context characteristics. Instrument selection was largely restricted to instruments validated by previous research in the relevant research fields (see section 2 for details).

The documentation of scales pertaining to the AES survey was first published along with the AES data in 2017 (Sacchi & Oesch, 2017).[2] The present documentation covers the extended, more complex database of the TREE2 baseline survey, which also includes data from the extension survey described above. From a methodological point of view, this raises the issue of potential survey-mode and setting effects: The AES assessment was conducted in a uniform proctored classroom setting supervised by carefully instructed test administrators; the extension survey, by contrast, took place in an unproctored individual setting outside of school.

1 The survey is part of an overarching assessment scheme implemented by the Swiss Conference of Cantonal Ministers of Education (EDK) to test basic skills in key subject areas at various stages of compulsory education. For details, see www.icer.unibe.ch and http://uegk-schweiz.ch/).

2 See forsbase.unil.ch/project/study-public-overview/16165/0/.

Furthermore, the latter employed two sequentially applied survey modes (web survey and paper-and-pencil questionnaire). With regard to scaling, this incongruence requires that we have to carefully check for measurement invariance across survey settings and modes. Consequently, this documentation includes a number of relevant invariance tests and parameters for all scales that are based on data from the extension survey.

Beyond psychometric scales stricto sensu, this documentation also includes a number of item sum scores based on two or more single items. However, we have not included scores of test results and other types of composite variables.[3]

For all scales and composites drawing exclusively on data of the AES assessment survey, we report the previously calculated parameters (Sacchi & Oesch 2017) in the technical appendix of this documentation. In doing so, we provide TREE2 data users with an overview of all scales and composite variables available in the TREE2 baseline survey in one single document (see particularly section 2). The introductory text describing the methods of calculation and estimation used and the parameters reported in the technical appendix largely corresponds to the 2017 AES documentation (ibid.).

For each of the scales, we report estimates (i.e., scores) of the individual scale values for all participating students. In addition, our documentation aims at enabling data users to assess the scales' quality and measurement invariance (cf. particularly the technical appendix). Last but not least, our documentation ought to allow scholars to replicate, if they wish to do so, the calculation of models, tests and scale parameters and compare them with alternative specifications.

In the following sections, we first specify some relevant aspects of the TREE2 baseline survey's design (1), the selection and adaptation of the scales (2) as well as the statistical modelling and calculation of the scale values (3). Finally, we specify how the scale-specific results, reliability and quality checks were calculated and give some information on how to interpret them (4).

[3] As for the scales, the extension survey considerably enlarges the database on which these scores rely.

1 Survey Design and Database

The data of the AES survey were collected by means of a computer-based classroom survey among a random sample of approximately 22,000 students who were in their last year of lower secondary education (i.e., the 11th year[4] of compulsory schooling).[5] The survey included a comprehensive test of basic mathematical skills, along with a computer-assisted self-interview (CASI) of approximately 45 minutes. Among other things, the student questionnaire covered a broad selection of psychometric and other item-based measures, which are the subject of this documentation.

AES implemented a modular design with two different versions of the questionnaire, each of which were administered to a randomised split-half of the total sample.[6] The main building block of one version was the mathematics module, which mainly covered student, teacher and classroom characteristics relevant to the successful acquisition of mathematical skills during compulsory education and to related didactical and pedagogical research. The core of the second version was a student background module co-designed by TREE to collect information on a broad range of resources of the surveyed students, their families and the schools they were attending at the moment of the survey. This module was specifically developed for the TREE2 panel survey in order to measure, as comprehensively as possible, the starting conditions deemed to be relevant for the respondents' further education and labour-market careers and their life courses in general. Both questionnaire versions included a common core ('general questions') that was completed by all students participating in AES. The common core incorporated items that are of general interest for the research objectives of both modules.

Due to the modular design of the AES questionnaire, a substantial part of the questionnaire pertaining to TREE-relevant starting conditions of post-compulsory pathways was administered to only half of the AES sample (see *Figure 1*). In order to complete the missing items for the respondents to the other half (termed 'maths sample split' in Figure 1), TREE carried out an out-of-school 'extension' survey immediately after the AES survey. With a few exceptions, the questionnaire used for this survey was equivalent to that of the background module in the AES

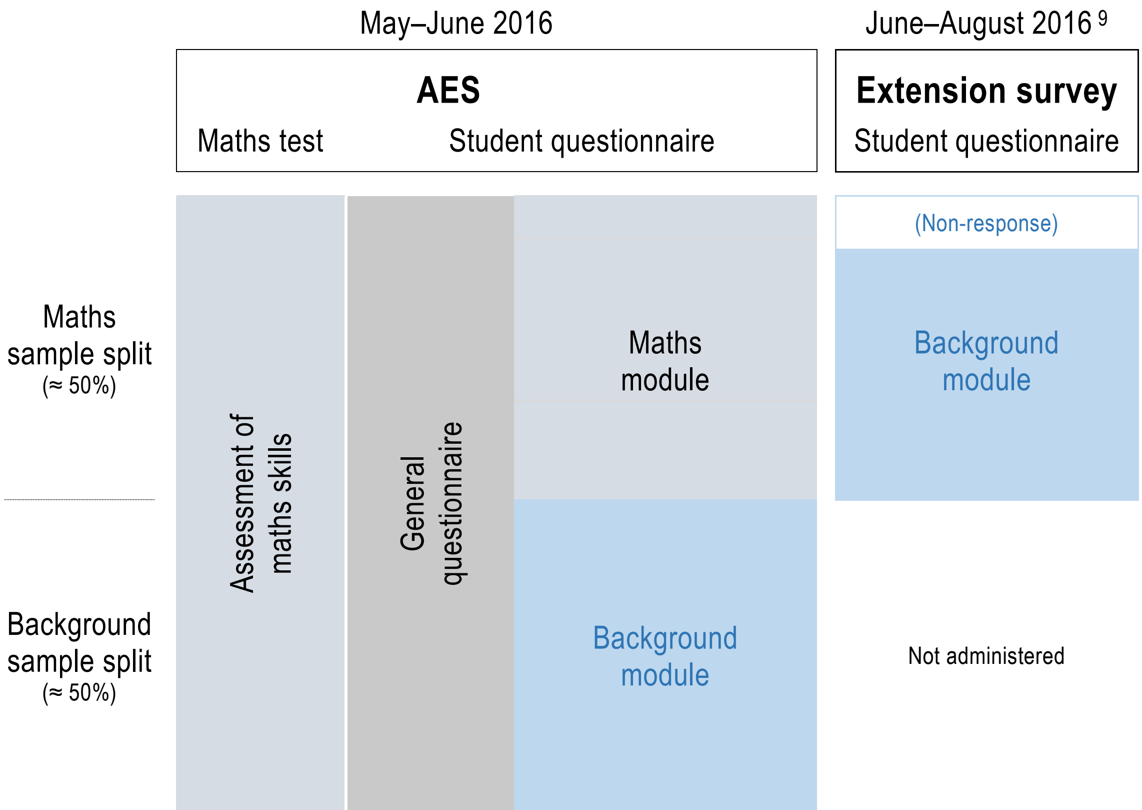
4 Including two years of kindergarten.

5 See Verner and Helbling (2019) for a detailed description of the sampling and the population.

6 The random assignment of the students to one questionnaire version was to guarantee that - within each school and each test session - both versions were evenly distributed over the 13 different test booklets used for the preceding mathematics assessment. Hence, from the students' perspective, booklet and questionnaire version were two independent, fully exogenous conditions.

survey, which was implemented in two 'standalone' versions, either in the form of a web or a paper-and-pencil questionnaire. The minor adaptations of the questionnaire under these changed setting and mode conditions included slightly modifying the order of instruments and adding a newly designed scale that had not been administered in the AES survey.[7] Apart from that, the web implementation was largely indistinguishable from the CASI instrument used by the AES.[8]

Figure 1: Design of the TREE2 baseline survey



In every canton, the extension survey was carried out as soon as the AES survey had been concluded in all sampled schools.[9] The web survey was implemented as the primary mode. Students who did not participate in the web survey received the questionnaire's paper-and-pencil version by mail as a secondary mode. As both survey modes are self-administered, they are well suited for the partly sensitive questionnaire items included in the extension survey. With this

[7] Two additional elements were placed at the end of the questionnaire: a brief cognitive skills test (KFT 4–12 + R; Heller & Perleth, 2000) as well as an experimentally varied repeated measurement of parental education.

[8] To maximise comparability with the AES CASI (and contrary to the web surveys in later TREE2 waves), the web mode was not adapted for smartphones (and respondents were asked to complete it on a computer).

[9] The median lag between the AES and extension survey was 29 days. 98 % of respondents completed the questionnaire between June and August, with a few pencil-and-paper questionnaires being returned up to the end of October.

mixed-mode design, the extension survey achieved a total response rate of almost 75% (73.3% if we consider only complete questionnaires; see also Table 1). Taking the relevant methodological literature into consideration, we do not expect significant mode effects (de Leeuw & Hox, 2011; de Leeuw, 2018; for proctored surveys see also Colosante et al., 2019).

As *Table 1* illustrates, the extension survey enabled us to substantially enlarge the available initial TREE2 sample base with a comprehensive measurement of relevant starting conditions. Among other things, this also allows for a more precise estimation of the scaling models and parameters that are at the centre of this documentation.[10] In light of the sample structure displayed in Table 1, it is important to address the issue of measurement invariance across the various survey settings and modes. That is why this documentation also provides statistical tests and quality measures that are relevant to this end (see section 4 and the technical appendix). The estimation of *setting effects* thereby draws exclusively on the CASI and the web survey, which rely on virtually interchangeable survey modes (i.e., it excludes the paper and pencil questionnaires, *n* = 15 608). And the estimation of *mode effects* draws exclusively on the extension survey (i.e., it excludes the classroom setting, *n* = 5 119). In doing so, we avoid the risk that the estimations of mode and setting effects are mutually confounded.

Table 1: Sample size and structure of the TREE2 baseline survey

	AES	Extension survey [1)]		Total
Survey Setting:	Proctored classroom survey	Unproctored individualised setting		
Survey Mode:	CASI	Web survey	P&P questionnaire	
(Sub-)sample size [2)]	11 124 [3)]	4 484	635	16 243

1) Including 89 incomplete questionnaires (with data for some scales only), which are treated as nonresponses when it comes to response statistics and the published sample weights (see also FN 10). 2) The number of cases for particular scales will generally be lower due to non-imputable missing values. 3) Background sample split (cf. Figure 1).

10 Regarding the scales partly relying on the extension survey, we draw on a customised sample weight tailored to the sample available for scaling purposes (cf. footer of Table 1). There are two types of non-negligible sample attrition, which exclusively affect the maths sample split (i.e., the unwillingness of AES respondents to provide their contact data for the TREE panel survey and non-participation in the extension survey). Given the high AES response rate of 93% (see Verner & Helbling, 2019: 39), the background split is therefore markedly less affected by attrition. The customised weight accounts for general and split-specific sources of attrition (see section 3.1.1a and FN 27 for further details).

These considerations do not affect the calculation of any of the scales administered in the general questionnaire and the AES maths module, as these scales do not rely on the extension survey. For calculations based on the general questionnaire, we can draw on data of the complete AES sample (approx. 22 000 students) and, for calculations based on the AES maths module, on the subsample to which the maths module was administered (approx. 11 000 students; cf. Figure 1). To ensure a statistically efficient estimate, the scaling models generally draw on the entire available sample base, including cases which, for various reasons, are not included in the scientific use files of the TREE2 dataset (Hupka-Brunner et al. 2023).[11]

In a survey administered in several languages, we also have to be careful regarding measurement invariance across survey languages (in our case German, French and Italian), which concerns all scales administered.[12] Basically, variance across languages can be the result of 'real' cultural or linguistic differences between language regions but also of inaccurate translations. That is why we report language-specific invariance tests and parameters (section 4 and appendix). As *Table 2* reveals, sample size substantially varies across survey languages.

Table 2: Breakdown of estimation samples by survey languages

Scales implemented in ...	General questionnaire	Background module	Math module
Available Estimation Sample [2)]	Full AES sample	Baseline survey [2)]	Math subsample
Survey Language:			
German	16 349	11 698	8 106
French	5 235	3 927	2 646
Italian	755	618	379

1) Number of cases for specific scales will in general be lower due to non-imputable missing values. 2) Cf. Table 1.

11 Data users who wish to estimate or replicate scaling models drawing on the complete database may do so. As the data excluded from the published data files are highly confidential, however, this is possible only on the premises of the study's headquarter in Bern and using a specially protected computer workplace.

12 In the AES, the survey language is identical with the teaching language of the sampled schools. In the extension survey, respondents were able to choose the survey language. In a few cases, this led to the situation that the extension survey was not completed in the same (national) language as the AES survey.

2 Selection and Adaptation of Scales

The AES questionnaire incorporated a broad range of more than 90 item-based instruments from relevant research areas (for theoretical considerations regarding the selection of instruments, see Hupka-Brunner et al. [2015] and Hascher et al. [2019]). As a general rule, preference was given to well-established, cross-disciplinary validated instruments used in surveys both in Switzerland and abroad.

A first selection of instruments was thoroughly pretested in the year preceding the main survey (2015).[13] One important objective of the pretest was to assess measurement properties of the preliminary selection of questionnaire instruments and scales in the Swiss context. This included assessments of the dimensionality, reliability and the cross-language measurement invariance of the scales. Some of the scales had to be newly translated to make them available in all survey languages. In these cases, the pretest was used to check measurement invariance across language versions and to improve improper translations. Moreover, the pretest was used to clean up scales with dodgy items, to shorten others and, lastly, to narrow down and optimise the selection of instruments for the main survey. We shortened many scales to three or four items to ensure a comprehensive coverage of relevant concepts without unduly increasing response burden and interview duration.

Wherever possible, the original instruments were implemented without modification in order to preserve measurement properties of the selected scales and to maximise data comparability. However, given the multitude of aspects to be considered in questionnaire construction (Dillman, Smyth & Christian, 2014), slight adaptations of the original instruments often could not be avoided.[14]

---

[13] The main objective of the pretest was to improve the assessment of mathematical skills, the design of the student questionnaire and the fieldwork for the main survey. The pretest sample was split evenly across the three test languages, German, French and Italian, and included more than 2 000 students from 70 schools.

[14] The manifold methodological, empirical and substantive reasons for such adaptations include the following: At the methodological level, there was the need to adapt instruments that were originally developed for a different survey mode (de Leeuw, Hox & Dillman, 2008: 311f.) and to standardise the format of each type of question in order to reduce the response burden and improve comprehensibility (Dillman, Smyth & Christian, 2014: 210f.). Empirically, the pretest in some instances uncovered insufficient cross-language measurement invariance, which suggested the need to check and, in some cases, improve the translations of the instruments. Finally, there was the requirement to closely replicate some of the instruments from the first TREE cohort (TREE1).

The modifications of the original instruments can pertain to both the question format and wording of stimuli as well as to the response scales and sometimes even to the items. In most cases, however, they are minor so that a substantial impact on the measurement properties and comparability of the resulting scales seems unlikely. It should also be noted that, for similar reasons, many popular scales are far less standardised in survey practice than generally perceived. Moreover, in the case of several circulating scale versions, the original version of the scale is not necessarily the most appropriate.

Table 3 conveys a topically ordered overview of all scales and item-based instruments that were implemented in the AES main field. The '*Positive Attitude towards Life*' scale was administered in the extension survey only. In a few cases, several scales partly rely on the same items. Consequently, they should not be introduced in one and the same multivariate model. Apart from scales involving main and sub-dimensions, the scales in question are framed by a dotted line in Table 3. For the *'Global self-esteem scale'* (and one of its subdimensions) a shortened version implemented in later waves of TREE2 is also available (see scale reporting in the appendix).

To enable comparative analyses between TREE1 and TREE2, the range of implemented instruments also includes some original scales used in the PISA 2000 survey, the baseline survey of the first TREE cohort (TREE1). For some of these scales (family wealth, social and cultural communication within the family), we implemented both the original version already used in PISA 2000 and an adapted version that was optimised for TREE2. The former is preferable for comparative analyses of both cohorts, the latter for analyses of the second cohort only.

Table 3: Item-based scales and composites (without scales for subdimensions)

Survey topic Scale / composite	[Variable name] 1)	AES questionnaire module 2)	Source 3)
Family background			
Family climate			
Emotional closeness to parents	[closep_comp]	Background	TREE1 - based on Szydlik, 2008
Parental pressure to achieve	[press_fs]	Background	Böhm-Kasper et al., 2000
Parents' achievement expectations	[expectp_fs]	Math	Hascher et al., 2019
Mother's achievement expectations	[expectm_fs]	Math	Hascher et al., 2019
Father's achievement expectations	[expectf_fs]	Math	Hascher et al., 2019
Mother's social norms about mathematics	[socnormsm_fs]	Math	PISA 2012
Father's social norms about mathematics	[socnormsf_fs]	Math	PISA 2012
Family educational support (PISA2000) 4)	[famedsup_fs]	Background	PISA 2000
Social communication (PISA2000) 4)	[soccom_fs]	Background	PISA 2000
Social communication (adapted TREE2)	[soccom_m_fs]	Background	PISA 2000 (adapted TREE2)
Social, cultural & economic resources			
Social capital (own)			
Perceived social network support	[closupp_fs]	Background	TREE2 (BHPS, ISSP 2003)
Cultural capital (family of origin)			
Parents: reading interest	[joyreadp_comp]	Background	TREE2
Cultural communication (PISA2000) 4)	[cultcom_fs]	Background	PISA 2000
Cultural communication (adapted TREE2)	[cultcom_m_fs]	Background	PISA 2000 (adapted TREE2)
Household possessions: classical culture (PISA2000) 4)	[cultposs_fs]	Background	PISA 2000
Cultural capital (own)			
Embodied cultural capital	[inccap_fs]	Background	TREE2
Cultural activities 5)	[cult_fs]	Background	PISA 2000 (partially adapted)

1) Student score variable names from 2023 TREE2 data release. 2) Database by module: General → full AES sample; background module → TREE2 baseline sample; math module → AES math sample split. 3) See technical appendix for a detailed list of sources. 4) Scales administered in the the first TREE cohort (TREE1). 5) A subscale of this scale has been adopted as is from PISA 2000 / TREE1 (cf. Table 4).

Table 3 (continued): Item-bases scales and composites

Survey topic Scale or composite	[Variable name] 1)	AES questionnaire module 2)	Source 3)
Social, cultural & economic resources (continued)			
Economic capital (family of origin)			
Household possessions: family wealth (PISA2000) 4)	[wealth_fs]	Background	PISA 2000
Household possessions: family wealth (adapted TREE2)	[wealth_m_fs]	Background	PISA 2000 (adapted TREE2)
Family affluence scale (FASIII)	[fasiii_comp]	Background	Hobza et al., 2017
Satisfaction and well-being			
Satisfaction			
Capabilities	[cap_fs]	Background	Sen, 1985; Anand & van Hees, 2006
School-related well-being			
Positive attitude towards school	[posatt_fs]	General	Hascher, 2004
Enjoyment in school	[enjoyschool_fs]	General	Hascher, 2004
Physical complaints in school	[physpain_fs]	General	Hascher, 2004
Worries about school	[trouschool_fs]	General	Hascher, 2004
Social problems in school	[socprob_fs]	General	Hascher, 2004
School reluctance	[schoolav_fs]	General	Hagenauer & Hascher, 2012 (modified)
Non-cognitive factors			
Motivational concepts			
Intrinsic achievement motivation	[achmoti_fs]	General	IGLU 2001
Extrinsic achievement motivation	[achmote_fs]	General	IGLU 2001
Instrumental learning motivation (PISA2000) 4)	[insmot_fs]	General	PISA 2000
Interest in reading (PISA2000) 4)	[intrea_fs]	General	PISA 2000
ICT interest	[ictintr_fs]	Math	ICILS 2013
Dispositional interest	[intsubj_fs]	Math	COACTIV 2008
Identified motivation (mathematics)	[instrumot_fs]	Math	PISA 2012
External motivation regulation	[extreg_fs]	Math	Ryan & Conell, 1989
Classroom participation	[engage_fs]	Math	Eder, 1995, 2007
Performance-approach goals (SELLMO)	[approxgoals_fs]	Math	SELLMO 2012
Learning goal orientation (SELLMO)	[learntarget_fs]	Math	SELLMO 2012
Work avoidance (SELLMO)	[avoidwork_fs]	Math	SELLMO 2012
Avoidance performance goals (SELLMO)	[avoidblame_fs]	Math	SELLMO 2012
Self-perception			
Global self-esteem 6)	[sel_fs]	Background	Rosenberg, 1979
General perceived self-efficacy scale (GSES)	[seef_fs]	Background	GSES (adapted TREE1)
Academic self-efficacy	[acaself_fs]	General	Hascher, 2004
Academic self-concept (PISA2000) 4)	[scacad_fs]	General	PISA 2000
Verbal self-concept (PISA2000) 4)	[scverb_fs]	General	PISA 2000
Maths self-concept	[matcon_fs]	General	PISA 2000 (adapted AES)
ICT self-concept	[ictabil_fs]	Math	ICILS 2013
Specific self-efficacy: numeracy	[selfeffa_fs]	(General) 7)	PISA 2012; Girnat, 2018
Specific self-efficacy: algebra	[selfeffb_fs]	(General) 7)	PISA 2012; Girnat, 2018
Specific self-efficacy: geometry	[selfeffc_fs]	(General) 7)	Girnat, 2018
Specific self-efficacy: probability	[selfeffd_fs]	(General) 7)	Girnat, 2018

1) Student score variable names from 2023 TREE2 data release. 2) Database by module: General → full AES sample; background module → TREE2 baseline sample; math module → AES math sample split. 3) See technical appendix for a detailed list of sources. 4) Scales administered in the surveys of the first TREE cohort (TREE1). 6) Data and scale appendix also include a shortened 7-item-version of this scale. 7) Half of the items implemented in the math module.

Table 3 (continued): Item-bases scales and composites

Survey topic Scale or composite	[Variable name] 1)	AES questionnaire module 2)	Source 3)
Non-cognitive factors (continued)			
Emotions related to maths classes			
Mathematics anxiety	[anxmath_fs]	Math	PISA 2012
Mathematics boredom	[boredom_fs]	Math	AEQ-M (short-version)
Mathematics anger	[anger_fs]	Math	AEQ-M (short-version)
Mathematics enjoyment	[enjoymath_fs]	Math	AEQ-M (short-version)
Volitional strategies			
Perseverance	[persev_fs]	General	PISA 2012
Effort: learning (PISA2000) 4)	[effper_comp]	Background	PISA2000
Personality characteristics			
Big five: extraversion	[big5_e_comp]	Background	Rammstedt et al., 2014
Big five: agreeableness	[big5_a_comp]	Background	Rammstedt et al., 2014
Big five: conscientiousness	[big5_c_comp]	Background	Rammstedt et al., 2014
Big five: neuroticism	[big5_n_comp]	Background	Rammstedt et al., 2014
Big five: openness	[big5_o_comp]	Background	Rammstedt et al., 2014
Internal locus of control	[loci_comp]	Background	GESIS (short version)
External locus of control	[loce_comp]	Background	GESIS (short version)
Values & attitudes			
Work-related extrinsic value	[vawe_fs]	Background	TREE1 - based on Watermann, 2000
Work-related intrinsic value	[vawi_fs]	Background	TREE1 - based on Watermann, 2000
Family value	[vafa_comp]	Background	TREE1
Positive attitude towards life	[posl_fs]	Extension survey	TREE1; Grob et al., 1991
Attitudes related to mathematics classes			
Reality-based learning	[realref_fs]	Math	Girnat, 2015, 2017
Discovery / exploratory learning	[disclearn_fs]	Math	Girnat, 2015, 2017
Social learning	[soccomlearn_fs]	Math	Girnat, 2015, 2017
Instructivist learning	[instreplearn_fs]	Math	Girnat, 2015, 2017
System aspect	[sysformasp_fs]	Math	Girnat, 2015, 2017
Scheme aspect	[schemasp_fs]	Math	Girnat, 2015, 2017
Application aspect	[applyasp_fs]	Math	Girnat, 2015, 2017
Education and training			
Characteristics of maths lessons (end of lower secondary education)			
Teacher: cognitive activation	[cogself_fs]	Math	COACTIV 2008
Teacher: classroom management	[classman_fs]	Math	COACTIV 2008
Teacher: individual learning support	[indsup_fs]	Math	COACTIV 2008
Teacher: instruction quality	[instqual_fs]	Math	PISA 2006
Situational interest	[intsit_fs]	Math	COACTIV 2008
Perceived autonomy support	[persuppauto_fs]	Math	Seidel, Prenzel & Kobarg, 2005
Perceived competence support	[persuppcomp_fs]	Math	Seidel, Prenzel & Kobarg, 2005
Perceived social relatedness	[persocincl_fs]	Math	Seidel, Prenzel & Kobarg, 2005
Classmates' appreciation of mathematics	[apprmath_fs]	Math	PISA 2012
Absenteeism / intention to change education			
Absenteeism / truancy	[truancy_fs]	General	PISA 2000, PISA 2012

1) Student score variable names from 2023 TREE2 data release. 2) Database by module: General → full AES sample; background module → TREE2 baseline sample; math module → AES math sample split. 3) See technical appendix for a detailed list of sources. 4) Scales administered in the first TREE cohort (TREE1).

In principle, all scales listed in Table 3 are one-dimensional, that is, they have been designed to measure *one* theoretical construct or latent dimension each.[15] However, some of the scales are composed of several sub-dimensions, each representing a facet of one overarching construct. As researchers may wish to distinguish between the sub-dimensions of these scales, the scientific use files of TREE2 also include student scores for each sub-dimension. The following table lists both the main and sub-dimensions of the scales in question.

Table 4 Scales with sub-dimensions

Scale – main dimension	Variable name 1)	Subdimensions	Variable name 1)
Background module scales			
Global self-esteem 2) 3)	[sel_fs]	Positive global self-esteem 4)	[sele_fs]
		Negative global self-esteem / depression 4) 5)	[seld_fs]
Embodied cultural capital	[inccap_fs]	Embodied cultural capital: manners	[manners_fs]
		Embodied cultural capital: verbal skills	[verbskill_fs]
Cultural activities	[cult_fs]	"Lowbrow" cultural activities	[cultlow_fs]
		"Highbrow" cultural activities (PISA2000) 6)	[culthigh_fs]
Math module scales			
Parents' achievement expectations	[expectp_fs]	Mother's achievement expectations	[expectm_fs]
		Father's achievement expectations	[expectf_fs]
Instructivist learning	[instreplearn_fs]	Instructivist learning: teachers' instructions	[instrlearn_fs]
		Instructivist learning: repetitive practice	[replearn_fs]
Social learning	[soccomlearn_fs]	Social learning: social arrangement	[soclearn_fs]
		Social learning: communication	[comlearn_fs]
System aspect	[sysformasp_fs]	System aspect: logical thinking	[systasp_fs]
		System aspect: formalism	[formasp_fs]
Teacher: cognitive activation 7)	[cogself_fs]	Cogn. activation: finding solutions & arguing	[cogselfa_fs]
		Cogn. activation: strategies and learning from mistakes	[cogselfb_fs]

1) The short names of the student score variables in the TREE2 scientific use file are given in brackets. 2) In accordance with Huang et al. (2012) and Donnellan et al. (2016), this scale is clearly two-dimensional in the TREE2 baseline survey. 3) Data and appendix also include a shortened 7-Item-Version of this scale (*sel_m_fs*). 4) Sub-dimension labels according to Huang et al. (2012). 5) Data and appendix also include a shortened 3-item-version of this subscale (*seld_m_fs*). 6) Corresponds to '*Cultactv*' scale in PISA 2000/TREE1. 7) As this scale is not one-dimensional in the AES survey, we distinguish two (inductively optimised) sub-dimensions.

Some of the instruments described in this documentation are based on two items only, making it impossible to fit any scaling model to the data. Henceforward, we call scores derived from

15 One should note, however, that the one-dimensionality of the selected scales may be empirically controversial. For one scale, *'Global Self-Esteem'* (according to Rosenberg, 1979; 2014), we are aware that this is the case (see von Collani & Herzberg, 2003; Huang & Dong, 2012; Donnellan, Ackerman & Brecheen, 2016). With respect to this scale, we decided to provide the student scores for both the one-dimensional model and for the two sub-dimensions described in the literature. Hence, we treat this scale the same way as other scales with sub-dimensions and leave it up to the data users to decide on the appropriate scaling solution.

mostly short, item-based instruments *item-based composites* (for an overview see *Table 5*).[16] In case of the *'Family affluence scale'* in Table 5, the term «scale» is a misnomer as it represents de facto a sum score, i.e., an item-based composite (for details, see Hobca et al., 2017).[17]

Table 5: Item-based composites

Concept [1)] / Dimension	Variable name [2)]	Number of items
Big Five Inventory		
Extraversion	[big5_e_comp]	2
Agreeableness	[big5_a_comp]	3 [3)]
Conscientiousness	[big5_c_comp]	2
Neuroticism	[big5_n_comp]	2
Openness	[big5_o_comp]	2
Locus of control		
Internal locus of control	[loci_comp]	2
External locus of control	[loce_comp]	2
Effort: learning (PISA2000) [4)]	[effper_comp]	2
Family values	[vafa_comp]	2
Parents: reading interest	[joyreadp_comp]	2
Emotional closeness to parents	[closep_comp]	2
Family affluence scale (FASIII) [FN17]	[fasiii_comp]	6

1) With the exception of 'Effort: learning' (general questionnaire, full sample), all composites belong to the background module. 2) The short variable names of the composite scores in the scientific use file are reported in brackets. 3) For the composite with one extra item, see Rammstedt and John (2007: 210). 4) This composite has been previously administered in the surveys of the first TREE cohort (TREE1).

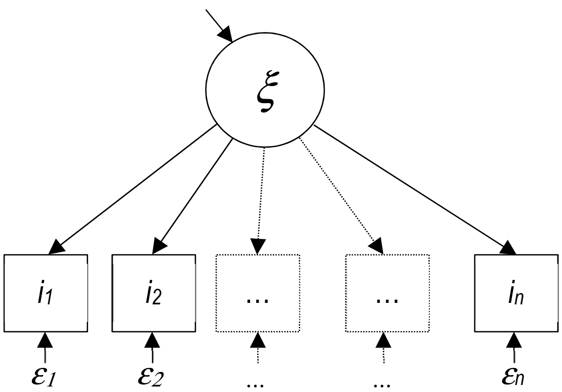
16 For item composites, student scores are calculated from imputed item ratings (cf. 3.1.1 b).

17 Note that this composite partly draws on the same items as the wealth scales in Table 3.

3 Statistical Modelling

As mentioned above, the scales in the AES questionnaire are item-based instruments intended to measure *one* theoretical construct each. Confirmatory factor analysis (CFA) is a common approach to the empirical estimation of latent (i.e., not directly observable) characteristics captured by such measurement instruments (see, e.g., Long, 1983; Schmitt, 2011). As our selection of scales is restricted to validated instruments that were designed to measure a common latent dimension, we limit ourselves to fitting a straightforward one-dimensional CFA model (see Aichholzer, 2017: 80–84) to each scale-specific item set. The CFA model illustrated in *Figure 2* relies on n items (i_1, i_2, ..., i_n) with associated item-level measurement errors ε_n, which all measure the same latent dimension ξ. For scales with several subdimensions (see Table 4 above), a separate CFA model is fitted to each subdimension.[18]

Figure 2: One-dimensional confirmatory factor model



For every model estimated hereafter, selected model parameters, fit statistics and scale quality measures are reported in the technical appendix (p. 34ff.). This includes a test of one-dimensionality, various measures of internal scale consistency as well as tests and indices of measurement invariance across survey languages and, where appropriate, survey settings and modes. Throughout this documentation, our primary focus is the quality of the scales (and the corresponding student scores) rather than model fit. If the fit of the straightforward one-factor model turns out to be poor, we neither modify the model to improve fit nor do we test alternative (e.g., multi-dimensional) models. It is up to the data user to judge whether the one-dimensional CFA models are appropriate and whether the scales have the required properties.

[18] An alternative approach would be to fit second-order CFA models to each dimension (Aichholzer, 2017: 89f.).

3.1 Estimation of the confirmatory factor models

In its standard form, structural equation modelling - including CFA as a special case - relies on a number of quite restrictive assumptions that are hardly ever met in practice. Basically, the observations should be independent, and the indicators should be measured on a continuous scale (interval-level measurement) and follow a multi-normal distribution (see, e.g., Hoyle, 2000). As regards the database of the AES and the TREE2 baseline survey, none of these assumptions holds: The two-stage sampling procedure implies that observations are clustered within schools (see Verner & Helbling, 2019) and hence are not independent. Moreover, measurement of the indicators is at ordinal (or binary) level as it mostly relies on Likert-type rating scales. And last but not least, the skewed univariate distributions of many ratings are hardly consistent with the required multivariate normality.

The methodological literature offers a wide range of suggestions on how to relax some of the assumptions of the standard SEM model and how to deal with ordinal, binary or skewed indicators and clustered observations (cf., e.g., Bryant & Jöreskog, 2016).[19] In particular, the suggestions include two-stage estimation methods that exploit polychoric correlations and generalised structural equation models (GSEM) that are suited for short response scales and categorical indicators (Rhemtulla, Brosseau-Liard & Savalei, 2012; Bryant & Jöreskog, 2016). However, there is currently no well-established, generally accepted estimation approach tailored to both ordinal indicators that are not normally distributed and a complex sample with clustered observations.

We therefore follow the recommendations of Rhemtulla et al. (2012; similarly Harpe, 2015: 843) regarding the accurate estimation of CFA models on the basis of ordinal, Likert-type indicators. They suggest two different estimation strategies depending on the length of the rating scales. For item responses that rely on a rating scale with at least five points (i.e., ordered discrete response categories), they suggest a two-step estimation based on polychoric correlations. For item evaluations that rely on shorter rating scales with four or less points, a generalised structural equation model (GSEM) is in order. Below, we describe these estimation strategies in more detail.[20] As our primary goal is to estimate accurate student scores, we also implement some

[19] Clustered observations may not only affect variance estimation and model fit but also bias the estimation of model parameters (i.e., factor loadings; cf. Stochl et al., 2016; Muthén & Satorra, 1995; Wu & Kwok, 2012).

[20] All calculations were performed using Stata version 15.0 (AES) and 16.1 (TREE2 baseline survey). For both strategies, model estimations in general converge without problems. In a few cases, mostly in multi-group models, it was necessary to constrain an error variance or to collapse smaller groups to achieve convergence (which is noted in the scale reporting of the scales concerned, see appendix).

sensitivity checks to assess the equivalence of student scores obtained via alternative model-estimation strategies (see section 3.2.1).

3.1.1 Two-step estimation based on polychoric inter-item correlations

The two-step approach starts with the estimation of a matrix of polychoric correlations between all items of a given scale (tetrachoric correlations, respectively, in the case of dichotomous items).[21] In the second step, maximum likelihood estimation is used to fit the one-dimensional CFA model from Figure 2 to the resulting correlation matrix.[22] The models are identified by setting the loading of the first item and the variance of the latent factor to one. The CFA models are also estimated separately for each of the three language subsamples. This allows for multi-group analysis designed to test and assess measurement invariance across the survey languages (see section 4 and, e.g., Steinmetz et al., 2008; Milfont & Fischer, 2015).

Below, we briefly describe how we deal with (a) the complex AES sample and (b) with missing item values in the context of the two-step estimation approach.

(a) Complex sample design and survey weighting

The AES survey relies on a random sample of students that was disproportionally stratified by cantons and type of cantonal curriculum (Verner & Helbling, 2019).[23] Furthermore, the samples analysed here are also affected by sample attrition. An unbiased estimation of any population characteristic therefore requires the *application of an appropriate survey weight* to account for the disproportional sampling design as well as for unit nonresponse. This also pertains to the estimation of polychoric correlations or the parameters of the CFA models to be estimated (e.g., factor loadings).[24]

---

[21] A polychoric correlation is defined as the maximum likelihood estimate of the correlation between two hypothetical, normally distributed continuous latent variables derived from two corresponding ordinal indicators. Estimations were calculated using the Stata package "polychoric" by Stas Kolenikov (from http://staskolenikov.net/stata).

[22] Maximum likelihood estimation has been found to be among the most appropriate estimation methods (together with ULS and DWLS; see Yang-Wallentin, Jöreskog & Luo, 2010) for analysing polychoric correlations derived from ordinal indicators.

[23] Lower secondary schools in Switzerland are mostly "tracked", that is, students are enrolled in separate programmes with varying academic requirements.

[24] Weighting would only be unnecessary in the case of a strict invariance of the postulated scaling model across subpopulations of any kind. If this strong assumption were met, the damage of unnecessarily applying survey weights would be limited to inflating the variances of the estimates to some degree (Bollen, Tueller & Oberski, 2013). Given the huge AES sample, this would not be too disturbing.

When estimating the polychoric correlations, we therefore use one out of three different survey weights, depending on whether a given scale is embedded in the background module, in the maths module or in the general questionnaire. For the scales from the latter two, we rely on the suitable AES weights.[25] With regard to AES, module-specific analyses require particular weights, as the sampling design of the randomised sample split for the distinct questionnaire modules (according to Figure 1) differs with respect to the shape of disproportional cantonal stratification.[26] On the basis of the module-specific AES weights, we have constructed an additional weight for the TREE2 baseline survey, which accounts not only for the AES sampling design and nonresponse but also for sample attrition in the extension survey.[27]

As regards the two-step estimation approach, it should be noted that variance estimation does not account for the clustering of observations within schools implied in the two-stage sampling (see Verner & Helbling, 2019).

(b) Handling of missing item values

Missing item values are not a major problem affecting the scales in the AES survey. As usual in surveys, however, there is a small share of missing item values, owing mainly to item non-response. With the exceptions mentioned below, the share of cases with missing information on at least one item of the scale does not exceed 5%. For two out of three scales, the percentage is below 1%.

A considerably higher share of missing values results for half of the items of each of the four scales that measure different facets of 'specific self-efficacy' in mathematics. This is a direct consequence of the questionnaire design (and therefore not a matter of methodological

25 We use the respective non-response adjusted weights from the AES scientific use file ('*smp_w_nrastubw*' for the scales of the general questionnaire and *'smp_w_qmath'* for the scales of the maths module).

26 The reason is that the design of the two complementary sample splits has been optimised for two different purposes: The sample split drawn for the background module is designed to maximise statistical power at the national level, whereas the maths module split is optimised for separate analyses of cantons. In a nutshell, this was achieved by developing a disproportional subsampling scheme that further reinforces the general overrepresentation of small cantons among the sample split with the maths module and reduces it among the sample split with the background module. The weights for the sample splits then correspond to the general survey weight from the AES scientific use file ('*smp_w_nrastubw*') multiplied by the inverse of the within-canton subsampling fraction (see also Verner & Helbling, 2019).

27 For the baseline survey, we use an entropy-balancing weight (cf. Hainmueller, 2012; Hainmueller & Xu, 2013) that compensates for the AES disproportionate sampling design (incl. non-response adjustments) and, as far as the math-sample split is concerned, for the non-response related to willingness to be (re-)contacted and to participate in the extension survey (for details, see the TREE2 documentation on weighting: Sacchi, forthcoming). For the purpose of scaling, the e-balancing weight for the TREE2 baseline survey was re-estimated by taking into account the somewhat looser definition of survey participation employed throughout the scaling process (see Table 1 and the explanatory text).

concern[28]), as half of the items of each of these scales were incorporated into the general questionnaire and the other half into the maths module. This implies that the share of missing item information is close to zero for the general questionnaire, whereas it rises to around 50% for the items implemented in the maths module.

A relatively high share of missing values is also observed for two measures in which students evaluate the items on a rating scale that includes an explicit 'don't know' option. This pertains to the scale measuring *'Perceived social network support'* (*closupp_fs*) and the two-item composite for *'Parental reading interest'* (*joyreadp_comp*). For both instruments, the share of missing information rises to 10.4 and 8.7%, respectively, when explicit don't-know answers are included.[29]

Finally, there are four instruments containing some items that could not be administered to a minor portion of the sample.[30] With one exception, the overall share of cases with at least one missing item does not exceed 5% in these instances.[31]

These special cases and exceptions notwithstanding, the fraction of missing items is low to very low for the bulk of the scales. Hence, the impact of missing item information is presumably limited.

We applied *multiple imputation* to cope with missing values when estimating the scaling models (Rubin, 1996; White, Royston & Wood, 2011). Basically, missing item information was imputed - scale-by-scale - on the basis of all valid items pertaining to the same scale. The imputed samples thus cover all cases with a valid response for at least one of the items of a given scale. Given the ordinal measurement level of the item ratings, we applied chained equations with an ordinal (or, in a few cases, binary) logit link to create samples with imputed values (Royston, 2011). Following the rules of thumb given in White et al. (2011: 388), we set the number of

---

28 The randomised allocation of students to questionnaire modules ensures that the missing-at-random assumption (MAR), which is crucial for the imputation of missing values, is almost perfectly met here.

29 Missing item values owing to explicit don't-know answers and item non-response were imputed together.

30 Some items referring to specific relatives (e.g., the father) have not been administered when the students previously indicated that these relatives do not exist (this pertains to the instruments: Family Education Support, Parents Achievement Expectations, Parents Reading Interest and Emotional Closeness to Parents). The resulting missing values were treated the same way as other types of missing information. Although this is perhaps not an ideal solution in these cases, a substantial bias seems unlikely given the mostly very low number of cases to which this applies.

31 The exception is the *'Family Educational Support'* scale (*famedsup_fs*) for which the share of cases with at least one missing item amounts to 14.6%. This owes mainly to the item tapping sibling support, which was not administered among students who previously indicated that they have no siblings (see footnote 29).

imputations to five.[32] For each imputed dataset, we separately calculated a matrix of polychoric correlations and combined it to estimate the CFA models.[33]

For each scale-specific CFA model, we calculated statistics and indices describing factor structures, model-fit and scale properties (see section 4 and the technical appendix).

3.1.2 Generalised structural equation model for short response scales

If scales rely on item evaluations with short response scales of four or less points (including binary items), they were analysed using a generalised structural equation model (GSEM), as recommended in the literature (Rhemtulla, Brosseau-Liard & Savalei, 2012; Bryant & Jöreskog, 2016). Model parameter estimates were derived in one step directly from the microdata through numeric integration.[34] Contrary to the two-step approach, this amounts to a full-information, true maximum likelihood method (Bryant & Jöreskog, 2016: 192). We henceforth adopted the GSEM version of a one-dimensional CFA model, mostly with an ordinal logit link to account for the ordinal measurement level of the item sets to be analysed.[35]

(a) Accounting for the complex survey design

GSEM, as implemented in Stata, is able to account for complex sample designs. In particular, we used survey weights (as described in 3.1.1a) to obtain unbiased population estimates of the model parameters and applied cluster-robust variance estimation, which controls for the clustering of students within schools. Still, we assume that there is no substantive variation in the measurement model across schools (cf. Wu & Kwok, 2012).

(b) Handling of missing item values

GSEM estimation proceeds on an equation-by-equation basis. In the context of a simple one-dimensional CFA model, this amounts to an implicit treatment (i.e., imputation) of missing item values, as each item is represented by a separate equation.

---

32 The relatively low number of imputations seems appropriate for two additional reasons: First, we are primarily interested in unbiased point estimates of population parameters (e.g., factor loadings) and to a lesser degree in between-imputation and sampling variances. Second, some exploratory reproducibility checks, as suggested by White et al. (2011: 387), indicate that the polychoric correlations and other point estimates are highly stable for an even smaller number of imputations.

33 After applying Fisher's z-transformation, we simply average the correlation matrices and transform them back (see also footnote 31).

34 Integration mostly relies on mean–variance Gauss–Hermite quadrature with seven integration points (StataCorp, 2017: 562).

35 The ordinal logit link reduces to a simple logit link for the two scales that include binary items.

One drawback of the GSEM approach is that the calculation of most established statistics to describe model fit and scale properties is not straightforward. This is why we complemented the GSEM estimations for the item sets with short response scales by a separately estimated two-step model, as described in section 3.1. If the resulting factor structures and student scores do not substantially differ from those obtained via the GSEM approach, this may be taken as indirect evidence that the two-step approach works sufficiently well and its assumptions are met (in the appendix, we therefore also check for the equivalence of both types of student scores). Hence, the model and scale statistics taken from the two-step CFA model are likely to be valid approximations as well.

3.2 Student scores

3.2.1 Calculation and robustness of student scores

For instruments relying on item rating scales of 5 or more points, the student scores in the scientific use file (and the related descriptive statistics in the appendix) represent *regression factor scores* (see StataCorp, 2017: 582f. for details) from the two-step CFA models described in section 3.1.1. For scales based on item sets with short response scales (four or less categories), the student scores in the SUF are *empirical Bayes means* based on the GSEM models (ibid.: 566). The *variable names assigned to the student scores* in the scientific use file are composed of a prefix indicating the survey wave (e. g. *'t2'* in case of the 2nd follow-up survey), the root of the variable names of the involved items and the suffix '*_fs*', which is used as a marker for student score variables. The corresponding suffix for the item composites from Table 5 is '*_comp*'. The *variable labels assigned to the student scores* and item composites correspond to those contained in the scale-specific documentation in the appendix. For an unequivocal interpretation of the student scores in the TREE2 scientific use file, we recommend inspecting the factor loadings (see section 4). As a general rule, however, a high factor score will indicate that students score high on the latent dimension that is designated by the label of the student score variable.

For all scales, the model, scale and test statistics reported in the appendix rely on the two-step estimation approach described in section 3.1.1. This explicitly also applies to those instruments based on short response scales, where the student scores (and the related factor-score descriptives in the appendix) are derived from a GSEM model. We also check the calculation of student scores for robustness by reporting the shared variance of both types of student scores (from SEM and GSEM) as measured by the coefficient of determination (CD) (see appendix: Equivalence of Scores from Two-Step Approach). If their shared variance is close to 100% (i.e., CD approaches

1), one may safely conclude, first, that the different modelling strategies have a negligible impact on student scores and, second, that it also seems reasonable to take the various fit and scale statistics obtained from two-step estimation as good approximations. As documented scale by scale in the appendix, the coefficient of determination is indeed close to 1 for most scales (> .94 for 42 out of 48 involved scales). There are six exceptions, however, in which the shared variance is substantially lower (between 60 and 90%), thus indicating that some of the additional assumptions needed for the two-step model have probably been violated. This pertains to the scales measuring *'Absenteeism / truancy'* (*truancy_fs*), *'Family wealth'* as indicated by home possessions (both scale versions: *wealth_fs*, *wealth_m_fs*), *'Cultural activities'* including one of its subscales (*cult_fs*, *culthigh_fs*) and students' *'Maths self-concept'* (*matcon_fs*). For these scales, the model and scale statistics reported in the appendix should be interpreted with great caution, if at all. Still, this does not indicate that the student scores estimated via the GSEM approach are biased in any way.

For an additional robustness check for the student scores, we re-estimated the confirmatory factor models in s single step directly from the student microdata by using the MLMV method (StataCorp, 2017: 574). This allows us to control for the complex survey design through weighting and cluster-robust estimation and, at the same time, to implement an alternative full-information maximum-likelihood approach to account for missing item values.

Let us again look at the shared variances between the student scores obtained via the MLMV method and those via the two-step approach described in section 3.1.1 (see appendix: Equivalence of Scores from Robust MLMV).[36] With the exception of the aforementioned wealth scale (both scale versions), the shared variances uniformly exceed 96% (i.e., CD > .96) for all of the 87 scales in this documentation. This can again be taken as indirect evidence that the additional assumptions of the two-step approach regarding multivariate normal distributions and the measurement level are mostly met and, hence, that the statistics and indices derived from it are valid. To sum up, the robustness checks imply that with the few exceptions mentioned above, student-score estimates are very robust across the three different estimation methods recommended for the type of data analysed here.[37]

3.2.2 Inclusion of student scores in multivariate statistical models

Instead of using the scale-specific student scores, there are often good reasons to embed scale-specific CFA models into a more comprehensive structural equation model of substantive

[36] A disadvantage of this method is that many statistics to judge model fit and scale qualities are unavailable.

[37] This may be due to the fact that we analyse short, one-dimensional scales based on a large sample.

interest and to fit them all together in one step (cf., e.g., Aichholzer, 2017). It should be noted, however, that simultaneous estimation of both the measurement and the substantive part of a structural equation model is not necessarily always the best choice (cf. Devlieger & Rosseel, 2017): When one analyses a subsample of limited size, for instance, robust estimation of more complex models may be impossible. Moreover, even when the sample is large, misspecification bias in one part of a complex model may spread to other parts when they are fitted in a single step. A two-step approach employing previously estimated factor scores to investigate the substantive part of the model may have methodological merits in this respect (ibid.). This approach also has methodological drawbacks, however, basically because it implicitly treats factor scores as error-free measures of the latent dimensions to be analysed.[38] Some of the resulting problems, possible biases and correction methods are discussed, for example, by Croon (2002), Lu and Thomas (2008), Jin et al. (2016), and Devlieger and Rossel (2017).

[38] A random extraction of plausible values from the posterior distributions of the CFA models could be a quite obvious solution to this. However, contrary to skills assessment, this is an uncommon approach in the scaling of questionnaire items, possibly because of the reduced convenience this entails for data analysis.

4 Scale-specific reporting: Content and interpretation

In this section, we outline the various statistics, indices and quality measures reported in the scale appendix. For each scale (or subscale; cf. Table 4), this report includes two pages with a variety of scale-specific statistics. Below, we take the scale that measures *'Parental pressure to achieve'* as an example to illustrate the scope and interpretation of scale-specific results. *Figure 3* and *4,* respectively, display the two pages of results for this scale as they appear in the appendix. Each scale reporting is linked with the full list of scales available in the baseline survey, and vice versa (link in the lower right corner of Figure 3). Unless otherwise specified, all reported results refer to the two-step estimation of the CFA model according to Figure 2. However, the student-scores descriptives refer to the scores obtained from the GSEM model, as the 'press' items are rated on a four-point scale (see section 3.2.1). The header of each scale-specific results section includes the name of the scale that is also used to label the related student-score variable in the 2023 data release (TREE, 2023). Furthermore, the headers specify the sample basis on which the calculations for the respective scales draw (baseline survey sample[39], full AES sample or maths sample split).

The *model and fit statistics* reported include two likelihood-ratio tests as well as various common goodness-of-fit statistics, as discussed in the SEM literature (cf. Schreiber et al., 2006). The *likelihood-ratio tests* compare the current against the saturated model and the baseline model (basically postulating uncorrelated items), respectively. Ideally, we would expect a non-significant likelihood-ratio test of the current against the saturated model, which, for the reasons given above, is an unlikely result, however (see also van der Eijk & Rose, 2015). Moreover, for a well-fitting model, we expect the *comparative fit index* (*CFI*) and the *Tucker–Lewis index* (*TLI*) to approach 1, whereas the *root mean square error of approximation* (*RMSEA*) and the *standardised root mean squared residual* (*SRMR*) should be close to 0. Conventional cut-off criteria indicating a good fit between the hypothesised model and the observed data are ≥ .95 for CFI and TLI ≤ .06 for RMSEA and ≤ .08 for SRMR (see Hu & Bentler, 1999). Regarding Figure 3, one could tentatively conclude that the one-dimensional CFA model fits the '*Parental pressure to achieve'* scale sufficiently well, with some reservations regarding RMSEA and TLI, however. Two fit measures designed to compare different models, *Akaike's information criterion* (*AIC*) and the *Bayesian information criterion* (*BIC*), are also reported. They may serve as a point of reference if data users wish to fit alternative scaling models to the data. Finally, the *coefficient of determination* (*CD*) may be considered as an alternative measure of composite reliability (in

[39] That is, the combined sample composed of the background split-half sample of the AES and the AES extension survey.

the sense of internal consistency; cf. Bollen, 1989: 220f.), to be interpreted similarly to the reliability measures below.

Figure 3: Example of the reported scale-specific results (first results page)

Scale: Parental pressure to achieve — Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	462	2	.000
	Baseline vs. saturated	20063	6	.000
2)	**Root mean squared error (RMSEA)**			.122
	90% Confidence interval: lower bound			.113
	90% Confidence interval: upper bound			.131
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			142462
	Bayesian Information Criterion (BIC)			142554
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.977
	Tucker-Lewis Index (TLI)			.931
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.026
	Coefficient of determination (CD)			.816

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.811
(Cronbach's alpha = .751)	
McDonald's Omega	.811

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.95
factor 2	-.04
factor 3	-.09
factor 4	-.18

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
press1	0.69	0.01	0.68	0.70
press2	0.69	0.01	0.68	0.71
press3	0.78	0.00	0.77	0.79
press4	0.71	0.01	0.70	0.72

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
press1	2.2	1.0	1	4	15488
press2	3.0	0.9	1	4	15491
press3	3.0	0.8	1	4	15488
press4	2.8	0.9	1	4	15490

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
press1	1.66	-1.38	0.68	2.99
press2	1.79	-3.56	-1.79	0.80
press3	2.35	-5.01	-2.26	1.38
press4	1.84	-3.48	-1.23	1.53

List of scales (wave 0)

The output section to the right of the model-fit statistics presents the results on *scale reliability and dimensionality*. Among the various conceptualisations of measurement reliability discussed in the literature (e.g., Bollen, 1989), *internal scale consistency* is the most widely used in practical research. One important reason for this is certainly that internal consistency may be easily assessed without additional re-test or parallel measurements of the indicators. It should also be noted, however, that consistency measures avoid several conceptual drawbacks of possible alternatives (see Bollen, 1989: 209ff.). We report three alternative measures of internal scale consistency: *Cronbach's Alpha* is still the most widespread, although much criticised, consistency measure (ibid.: 217; Sijtsma, 2009; Revelle & Zinbarg, 2009; Trizano-Hermosilla & Alvarado, 2016). In a nutshell, it is widely recognised that alpha underestimates internal consistency if the indicators are ordinal or congeneric (i.e., not tau-equivalent) as is typical of most practical research situations. We nevertheless do report the classical version of alpha as it is part of most survey documentations and — if interpreted as a lower-bound estimate of internal scale consistency — may still be useful for comparative purposes.[40] In addition, we also report *Ordinal Cronbach's Alpha*, which is calculated the same way as classical alpha but from the matrix of polychoric instead of Pearson correlations (see Gadermann, Guhn & Zumbo, 2012: 5). This avoids downward bias owing to ordinal measurement. Finally, we also report *McDonald's Omega*, which is one of the most recommended measures of internal consistency. Omega is calculated on the basis of the factor loadings of the one-dimensional CFA model (according to formula 1 in Trizano-Hermosilla & Alvarado, 2016), which implies that it is adjusted for ordinal measurement. As omega is appropriate for congeneric indicators, it is probably the most adequate measure overall of internal scale consistency in our context (see also Yang & Green, 2015). Basically, values close to 1 indicate high internal consistency for all three measures. Looking at Figure 3, many researchers would probably interpret the identical ordinal alpha and omega values of .811 each as an indication of a 'good', consistent scale. It should be noted, however, that the widely used rules of thumb to determine whether internal scale consistency can be considered 'acceptable' or 'good' (usually values above .7 and .8, respectively) are not without problems. First, there exist various such rules of thumb with different critical thresholds. Second, and more importantly, such rules should not be applied blindly, as the acceptable level of internal consistency depends strongly on the type of analysis to be performed (Lance, Butts & Michels, 2006).[41]

A crucial assumption of the estimated CFA models is that the analysed item set captures only one latent construct. Therefore, we have also included a *test of the assumed one-dimensionality.* However, assessing dimensionality of Likert-type items is quite 'risky business', as van der Eijk

[40] The Stata package "Alphawgt", which allows for weights, was used to calculate alpha (Jann, 2004).

[41] There are some rather dubious rules of thumb that distinguish different levels of internal scale consistency (i.e., Cronbach's alpha). A popular variant is: $\alpha < .5$: unacceptable; $.5 \leq \alpha < .6$: poor; $.6 \leq \alpha < .7$: questionable; $.7 \leq \alpha < .8$: acceptable; $.8 \leq \alpha < .9$: good; $.9 \leq \alpha$: excellent
(cf. https://en.wikipedia.org/wiki/Internal_consistency, accessed on June 23, 2020).

and Rose (2015) put it. We used explorative factor analysis of polychoric correlations followed by Horn's parallel analysis to assess the dimensionality of the item sets, which proves to be a comparatively well-performing method (ibid.; Garrido, Abad & Ponsoda, 2013).[42] Basically, we applied an eigenvalue criterion that was corrected for random factors to account for sampling variance to determine the number of factors to be retained. In Figure 3, this approach gives us no reason to believe that the achievement-pressure scale is not one-dimensional, as only the eigenvalue of the first factor exceeds the critical value of zero. If we leave aside the scales composed of several sub-dimensions (cf. Table 4), the eigenvalues of the second factor are mostly below or only very slightly above zero for most of the scales in this documentation.[43] This being the case, we have no clear indication that the one-dimensionality assumption is violated.

The section below the model-fit statistics in Figure 3 documents the *standardised factor loadings* for each item, including standard errors and the confidence intervals. The item names correspond to those in the scientific use file (without the prefix-marker for the survey wave). High standardised loadings above, say, .6 or .7 indicate that neither measurement errors nor strong unique factors contribute excessively to the variance of the observed indicators. Almost all loadings reported in the appended scales reach this level. Occasionally, however, items show noticeably weaker loadings below .5 or even below .4, which some researchers may consider problematic. Eventually, the definition of an acceptable factor loading remains arbitrary and depends on the type of analysis, the number of scale items affected and the quality as well as the overall internal consistency of the scale (ibid.). As in other respects, we prefer to leave it to the data users to judge a particular scale's qualities.

To the right of the loadings, a number of *item descriptives* are reported, including the mean, the standard deviation, the range of the rating scale applied for item evaluation (min., max.) and the number of students with valid item data (see section 3.1.1b).

At the bottom of the first page of our scale-specific results, we report the *parameters of the categorical GSEM model* (cf. section 3.1.2) where it is estimated. Note that for this model, there are two types of item-specific parameters, namely, factor coefficients ('*coef*') that measure the effect of the latent variable on the indicator rating, and the estimated cut points ('*cutx*') on the logit distribution that separate the rating scale category 1 from category 2, category 2 from category 3 and so on. Hence, the number of estimated 'cut' parameters equals the number of ordered rating categories minus one. Remember that the GSEM model is used to generate

[42] The parallel analysis relies on the user-written "paran" package (Dinno, 2009).

[43] Exception: the two wealth scales.

student scores (see section 3.1) where students' item evaluations rely on short rating scales with four or less points (as documented by the item descriptives).

A second page of scale-specific results (see Figure 4 below) is dedicated to tests and indices that assess *measurement invariance across survey languages* and, where appropriate, *across survey settings and modes*. This is an important facet of measurement quality, as student scores obviously should be comparable – i.e., measure the same concepts on a possibly invariant scale – across all kinds of measurement conditions and subsamples of the underlying student population. We focus on some of the most crucial tests suggested in the literature on the multi-group analysis of measurement invariance (e.g., Vandenberg & Lance, 2000; Milfont & Fischer, 2015) to assess cross-language measurement equivalence. On top of the second results page, we first report a chi-square test of the *equality of the item-covariance matrices* across survey languages

Figure 4: Example of the reported scale-specific results (second results page)

Scale: Parental pressure to achieve (continued) Baseline survey sample

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	1717	28	.000	105	14	.000	26	14	.027

Tests of measurement invariance across …	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	31	6	.000	33	3	.000	11	3	.013
Strong invariance (plus equal intercepts)	923	6	.000	11	3	.010	4	3	.317
Strict invariance (plus equal error variances)	73	6	.000	12	3	.008	3	3	.413

Configural factor similarity across …	**Survey languages**		**Survey settings**		**Survey modes**	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.999	classroom vs. unproctored	.998	web vs. PAP	.996
	French vs. Italian	.997				
	Italian vs. German	.993				

Factor score equivalence: group specific vs. invariant models for …	**Survey languages**		**Survey settings**		**Survey modes**	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	1.000	unproctored	.999	PAP	.990
	Italian	.980				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
press_fs	0.0	0.9	-2.4	1.7	15535

Share of cases with imputed missing values: 0.6%
(Equivalence of scores from robust MLMV: CD = .997)
(Equivalence of Scores from Two-Step-Approach: CD = .984)

(German, French, Italian; cf. Table 2) and, when a scale relies on the TREE2 baseline survey (including the AES extension survey), across survey settings (classroom vs. unproctored) and survey modes (web survey vs. paper-and-pencil questionnaire (PAP); cf. Table 1).[44] If the hypothesis of equal covariance matrices is not rejected, this would be a strong indication of measurement invariance, making any further tests obsolete (ibid.).

The chi-square tests assembled in the section below refer to the one-dimensional CFA model from section 3.1.1, which was re-estimated separately for each survey language and, where appropriate, for each survey setting and survey mode. Hence, the tests assume that a common latent dimension exists, and its invariance is investigated by means of multi-group analysis. The three tests are designed to distinguish different levels of measurement equivalence, as discussed in the literature (ibid.). The first test is for *metric measurement invariance*, that is, for equal factor loadings. A non-significant test indicates that there is no evidence against the postulated invariance of the factor loadings across the different survey conditions. The second test takes the model with invariant loadings as its baseline and tests it against an alternative model with invariant loadings *and* intercepts, which implies *strong measurement invariance*. Third and lastly, the latter model is tested against an alternative positing *strict measurement invariance*, which furthermore requires invariant error variances (ε_i in Figure 2). Given the nested structure of the compared models, strong invariance would require that the first two tests be not significant and strict invariance that all three tests be not significant. Although this is a rather standard approach to assess measurement equivalence, the reservations against chi-square-based fit statistics discussed above also extend to chi-square-based multi-group comparisons: Even if the cross-language variations in the model parameters are negligible, these tests will nearly always be significant given the mostly huge samples analysed here. That is to say, a level of measurement equivalence that would be adequate for nearly all practical research purposes would still not be enough to pass these tests. Against this background, it is rather surprising that, with regard to the *'Parental pressure to achieve'* scale (see Figure 4), *strong or even strict measurement invariance* is not rejected ($p < .01$) with respect to survey modes (where, however, the test samples are smaller than for survey languages or settings; cf. Table 1).

Below the section with the chi-square-based invariance tests, we report two additional measures of factor equivalence, which will perhaps do better in meeting the practical needs of many data analysts. The first one, *Tucker's congruence coefficient* (*TCC*), is a measure of *configural factor invariance* (calculated according to formula 1 in Lorenzo-Seva & ten Berge, 2006). Basically, it

[44] Technically, this was achieved by specifying a multi-group model without a latent dimension and then testing a completely unconstrained model against a constrained one with equal variances and inter-item covariances.

is a pattern-similarity measure that approaches 1 when the loading patterns observed in two groups or conditions are identical. We report the coefficient separately for each pair of survey languages as well as for the pairs of survey settings and survey modes, where appropriate. According to Lorenzo-Seva and ten Bergen (ibid.: 61), two factors may be considered as approximately equal for practical purposes if TCC is .95 or higher. If we look at the scales documented in the appendix, this criterion is met for all pairwise comparisons across survey languages, survey settings and survey modes.

In addition, we also assess the degree of *micro-level factor equivalence at the level of student scores*. For this, we compare the student scores taken from an unconstrained model fitted separately for each language, setting or mode, respectively, with the student scores taken from a model for the entire sample on the assumption of strong measurement invariance (i.e., equal loadings and intercepts). If the differences between the former and latter are negligible across the analysed survey conditions, this is a strong indication that – from a practical point of view – the measurement can be regarded as sufficiently invariant. As a measure of micro-level agreement, we report – separately for each of the subsamples delineated by survey language, survey setting, and survey mode – the *coefficient of determination* (*CD*), which is calculated by regressing the student scores from the strong-invariance model on those from the unconstrained condition-specific models. Where the CD indicates that both scores share, say, 98% of their variance (i.e., CD ≥ .98), deviations from the postulated strong invariance model may be regarded as negligible. All scales in the appendix satisfy this criterion with respect to mode and setting effects. With regard to survey languages, there are some differences in a limited number of cases, which mostly concern the Italian language. It should be noted, however, that a perfect agreement cannot always be expected even if the 'true' measurement model was absolutely invariant as the estimated student scores also include some random error. This is particularly true for the scores gained through the separate analysis of small subsamples, as is the case for the Italian questionnaire (n = 379 – 755, cf. Table 2) and the paper-and-pencil mode (n = 635; cf. Table 1) of the extension survey (cf. Figure 1). Notably for these subsamples, the sampling errors in the factor loadings and hence also in the student scores are likely to be more substantial.[45] With this in mind, one could also accept a coefficient of determination of, say, .95 as an indication of a still fair level of measurement equivalence. Also with regard to language-specific invariance, almost

[45] In combination with skewed item distributions, this is probably also the reason why a few of the models underlying the invariance tests did not converge so that the subsamples for the French and the Italian languages had to be collapsed for this purpose. We added an explanatory note at the end of the measurement-equivalence section in the appendix, which is shaded in grey in these cases (e.g., the *'School reluctance'* scale).

all scales in the appendix satisfy this criterion.[46] In the case of the achievement-pressure scale in Figure 4, however, our results are unambiguous and do suggest a high degree of measurement equivalence across survey languages, settings and modes.

In the section following the measurement invariance tests and indices, we report the short variable names (*press_fs* in Figure 3) of the student score variables in the scientific use file (from either ML-SEM or GSEM, depending on the length of the rating scales; see section 3.2.1).[47] The respective descriptive statistics refer to the sample base used for the calculation of the student scores (including cases not published in the scientific use files of the data release; cf. section 1).[48]

Either one or two measures of factor-score equivalence across different estimation methods are reported at the bottom of the second results page (see section 3.2.1), depending on the length of the rating scales applied for item evaluation. With regard to the achievement-pressure scale in Figure 4, they confirm a high degree of equivalence between the student scores from all three estimation procedures.

[46] Exceptions to the rule: the Italian versions of '*vawe*', '*ictintr*', '*cogselfb*' and '*cultposs*'. In the case of '*cultposs*', this applies to the French version as well.

[47] The full variable names include an additional prefix to distinguish TREE2 survey waves (e.g., "t2" for the second follow-up survey).

[48] Relevant sample sizes are reported under "Factor score descriptives: Obs.". We also report the share of cases with imputed item values.

References

Aichholzer, Julian (2017). *Einführung in lineare Strukturgleichungsmodelle mit Stata*. Springer: Wiesbaden (1. Auflage).

Bollen, Kenneth A. (1989). *Structural Equation with Latent Variables*. Wiley: New York.

Bollen, Kenneth, Stephen Tueller & Daniel Oberski (2013). 'Issues in the Structural Equation Modeling of Complex Survey Data.' International Statistical Institute (Eds.): *Proceedings of the 59th World Statistics Congress*: Hongkong.

Bryant, Fred B. & Karl G. Jöreskog (2016). 'Confirmatory Factor Analysis of Ordinal Data Using Full-Information Adaptive Quadrature.' *Australian & New Zealand Journal of Statistics* 58 (2): 173–196.

Colasante, Emanuela, Elisa Benedetti, Loredana Fortunato, Marco Scalese et al. (2019). 'Paper-and-pencil versus computerized administration mode: Comparison of data quality and risk behavior prevalence estimates in the European school Survey Project on Alcohol and other Drugs (ESPAD).' *PLOS ONE* 14 (11): e0225140.

Croon, Marcel A. (2002). 'Using Predicted Latent Scores in General Latent Structure Models', pp. 195–224, in George A. Marcoulides & Irini Moustaki (Eds.): Latent Variable and Latent Structure Models. Lawrence Erlbaum: Mahwah, New Jersey.

de Leeuw, Edith D. (2018). 'Mixed-Mode: Past, Present, and Future.' *Survey Research Methods* 12 (2). 75–89.

de Leeuw, Edith, Joop Hox & Don. A. Dillman (2008). 'Mixed-Methods Surveys: When and Why', pp. 299–316, in Edith de Leeuw, Joop Hox & Don A. Dillman: *International Handbook of Survey Methodology.* Lawrence Erlbaum: New York.

de Leeuw, Edith & Joop Hox (2011). 'Internet Surveys as Part of a Mixed-Mode Design', pp. 45–76, in Marcel Das, Peter Ester & Lars Kaczmirek (Eds.): *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies.* Routledge: New York.

Devlieger, Ines & Yves Rosseel (2017). 'Factor Score Path Analysis.' *Methodology* 13 (Supplement 1): 31–38.

Dillman, Don. A., Jolene D. Smyth & Leah Melani Christian (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons: Hoboken (4th Edition).

Dinno, Alexis (2009). 'Implementing Horn's Parallel Analysis for Principal Component Analysis and Factor Analysis.' *Stata Journal* 9 (2): 291–298.

Donnellan, M. Brent, Robert A. Ackerman & Courtney Brecheen (2016). 'Extending Structural Analyses of the Rosenberg Self-Esteem Scale to Consider Criterion-Related Validity: Can Composite Self-Esteem Scores Be Good Enough?' *Journal of Personality Assessment* 98 (2): 169–177.

Gadermann, Anne M., Martin Guhn & Bruno D. Zumbo (2012). 'Estimating Ordinal Reliability for Likert-Type and Ordinal Item Response Data: A Conceptual, Empirical, and Practical Guide.' *Practical Assessment, Research & Evaluation* 17 (3): 1–13.

Garrido, Luis Eduardo, Francisco José Abad & Vicente Ponsoda (2013). 'A New Look at Horn's Parallel Analysis with Ordinal Variables.' *Psychological Methods* 18 (4): 454–474.

Gnambs, Timo & Kai Kaspar (2016). 'Socially Desirable Responding in Web-Based Questionnaires: A Meta-Analytic Review of the Candor Hypothesis.' *Assessment* 24 (6): 746–762.

Hainmueller, Jens (2012). 'Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies.' *Political Analysis* 20 (1): 25–46.

Hainmueller, Jens & Yiqing Xu (2013). 'ebalance: A Stata Package for Entropy Balancing.' *Journal of Statistical Software* 54 (7): 18.

Hair, Joseph F., William C. Black, Barry J. Babin, Rolph E. Anderson et al. (2006 [1987). *Multivariate Data Analysis* (6 ed.). New Jersey: Pearson Prentice Hall.

Harpe, Spencer E. (2015). 'How to Analyze Likert and Other Rating Scale Data.' *Currents in Pharmacy Teaching and Learning* 7 (6): 836–850.

Hascher, Tina, Christian Brühwiler, Andrea Erzinger, Boris Girnat et al. (2019). *Erläuterungen zu den Skalen des Kontextfragebogens Mathematikteil: Theoretischer Hintergrund und Forschungsinteressen.* Universität Bern, Pädagogische Hochschule St. Gallen & Pädagogische Hochschule FHNW (Eds.): Bern.

Hobza, V., Hamrik, Z., Bucksch, J., & De Clercq, B. (2017). The Family Affluence Scale as an Indicator for Socioeconomic Status: Validation on Regional Income Differences in the Czech Republic. *International journal of environmental research and public health*, 14(12), 1540-1549. doi: 10.3390/ijerph14121540.

Hoyle, Rick H. (2000). 'Confirmatory Factor Analysis', pp. 465–497, in Steven D. Brown (Ed.): *Handbook of Applied Multivariate Statistics and Mathematical Modeling.* Academic Press: San Diego.

Hu, Litze & Peter M. Bentler (1999). 'Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives.' *Structural Equation Modeling* 6 (1): 1–55.

Huang, Chiungjung J. & Nianbo B. Dong (2012). 'Factor Structures of the Rosenberg Self-Esteem Scale: A Meta-Analysis of Pattern Matrices.' *European Journal of Psychological Assessment* 28 (2): 132–138.

Hupka-Brunner, Sandra, Ben Jann, Thomas Meyer, Christian Imdorf et al. (2015). *Erläuterungen zum Kontextfragebogen der ÜGK 2016: Allgemeiner Teil.* Universität Bern: Bern.

Hupka-Brunner, Sandra, Thomas Meyer, Stefan Sacchi, Ben Jann et al. (2023). *'TREE2 study design. Update 2023'.* TREE: Bern. URL: boris.unibe.ch/id/eprint/175367

Jann, Ben (2004). *ALPHAWGT: Stata Module to Compute Cronbach's Alpha for Weighted Data*. Statistical Software Components S4444101). Boston: Boston College Department of Economics.

Jin, Shaobo, Hao Luo & Fan Yang-Wallentin (2016). 'A Simulation Study of Polychoric Instrumental Variable Estimation in Structural Equation Models.' *Structural Equation Modeling* 23 (5): 680–694.

Lance, Charles E., Marcus M. Butts & Lawrence C. Michels (2006). 'The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say?' *Organizational Research Methods* 9 (2): 202–220.

Long, J. Scott (1983). *Confirmatory Factor Analysis*. Vol. 33 of the Sage University Paper Series on Quantitative Applications in the Social Science, edited by Michael S. Lewis-Beck. Sage: Newbury Park.

Lorenzo-Seva, Urbano & Jos M. F. ten Berge (2006). 'Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity.' *Methodology* 2 (2): 57–64.

Lu, Irene R. R. & D. Roland Thomas (2008). 'Avoiding and Correcting Bias in Score-Based Latent Variable Regression with Discrete Manifest Items.' *Structural Equation Modeling* 15 (3): 462–490.

Milfont, Taciano L & Ronald Fischer (2015). 'Testing Measurement Invariance across Groups: Applications in Cross-Cultural Research.' *International Journal of Psychological Research* 3 (1): 111–130.

Muthén, Bengt O. & Albert Satorra (1995). 'Complex Sample Data in Structural Equation Modeling.' *Sociological Methodology* 25: 267–316.

Rammstedt, Beatrice & Oliver P. John (2007). 'Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German.' *Journal of Research in Personality* 41 (1): 203–212.

Revelle, William & Richard E. Zinbarg (2009). 'Coefficients Alpha, Beta, Omega, and the GLB: Comments on Sijtsma.' *Psychometrika* 74 (1): 145–154.

Rhemtulla, Mijke , Patricia E. Brosseau-Liard & Victoria Savalei (2012). 'When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods under Suboptimal Conditions.' *Psychological Methods* 17 (3): 354–73.

Rosenberg, Morris (1979). *Conceiving the Self*. Basic Books: New York.

Rosenberg, Morris (2014). 'Self-Esteem Scale.' GESIS (Ed.): *Zusammenstellung sozialwissenschaftlicher Items und Skalen*: Mannheim. ZIS-Version 16.00. doi: 10.6102/zis46.

Royston, Patrick (2011). 'Multiple Imputation by Chained Equations (MICE): Implementation in Stata.' *Journal of Statistical Software* 45 (4): 1–20.

Rubin, Donald B. (1996). 'Multiple Imputation After 18+ Years.' *Journal of the American Statistical Association* 91 (434): 473–489.

Sacchi, Stefan & Dominique Oesch (2017). 'ÜGK 2016: *'Assessment of mathematics skills: Documentation of questionnaire-based scales'*. Assessment of the Achievement of Basic Educational Competences' TREE / University of Bern: Bern.

Sacchi, S., & Krebs-Oesch, D. (2023). *Documentation of scales implemented from panel wave 1 onwards*. In TREE (Ed.), Transitions from Education to Employment, Cohort 2 (TREE2), panel waves 0-3 (2016-2019) (2.0.0) [Dataset]. University of Bern. Distributed by FORS data service. https://doi.org/10.48573/kz0d-8p12

Schmitt, Thomas A. (2011). 'Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis.' *Journal of Psychoeducational Assessment* 29 (4): 304–321.

Schreiber, James B., Amaury Nora, Frances K. Stage, Elizabeth A. Barlow et al. (2006). 'Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review.' *The Journal of Educational Research* 99 (6): 323–338.

Sijtsma, Klaas (2009). 'On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha.' *Psychometrika* 74 (1): 107–120.

StataCorp (2017). *Stata Structural Eqation Modeling Reference Manual. Release 15.* College Station, Texas: Stata Press.

Steinmetz, Holger, Peter Schmidt, Andrea Tina-Booh, Siegrid Wieczorek et al. (2008). 'Testing Measurement Invariance Using Multigroup CFA: Differences between Educational Groups in Human Values Measurement.' *Quality & Quantity* 43 (4): 599–616.

Stochl, Jan, Peter B. Jones, Jesus Perez, Golam M. Khandaker et al. (2016). 'Effects of Ignoring Clustered Data Structure in Confirmatory Factor Analysis of Ordered Polytomous Items: A Simulation Study Based on PANSS.' *International Journal of Methods in Psychiatric Research* 25 (3): 205–219.

Tabachnick, Barbara G & Linda S Fidell L. S. (2013 [1996]). *Using multivariate statistics* (6. ed.). Boston, MA: Pearson.

TREE (2023), Transitions from Education to Employment, Cohort 2 (TREE2), panel waves 0-3 (2016-2019) (2.0.0) [Dataset]. University of Bern. Distributed by FORS data service. https://doi.org/10.48573/kz0d-8p12

Trizano-Hermosilla, Italo & Jesús M. Alvarado (2016). 'Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements.' *Frontiers in Psychology* 7. Published online 2016 May 26: doi: 10.3389/fpsyg.2016.00769.

van der Eijk, Cees & Jonathan Rose (2015). 'Risky Business: Factor Analysis of Survey Data – Assessing the Probability of Incorrect Dimensionalisation.' *PLoS ONE* 10 (3). https://doi.org/10.1371/journal.pone.0118900

Vandenberg, Robert J. & Charles E. Lance (2000). 'A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research.' *Organizational Research Methods* 3 (1): 4–70.

Verner, Martin & Laura A. Helbling (2019). *Sampling ÜGK 2016: Technischer Bericht zu Stichprobendesign, Gewichtung und Varianzschätzung bei der Überprüfung der des Erreichens der Grundkompetenzen 2016.* Institut für Bildungsevaluation (IBE): Zurich.

von Collani, Gernot & Philipp Yorck Herzberg (2003). 'Zur internen Struktur des globalen Selbstwertgefühls nach Rosenberg.' *Zeitschrift für Differentielle und Diagnostische Psychologie* 24 (1): 9–22.

White, Ian R., Patrick Royston & Angela M. Wood (2011). 'Multiple Imputation Using Chained Equations: Issues and Guidance for Practice.' *Statistics in Medicine* 30 (4): 377–399.

Wu, Jiun-Yu & Oi-man Kwok (2012). 'Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches.' *Structural Equation Modeling* 19 (1): 16–35.

Yang-Wallentin, Fan, Karl G. Jöreskog & Hao Luo (2010). 'Confirmatory Factor Analysis of Ordinal Variables with Misspecified Models.' *Structural Equation Modeling* 17 (3): 392–423.

Yang, Yanyun & Samuel B. Green (2015). 'Evaluation of Structural Equation Modeling Estimates of Reliability for Scales with Ordered Categorical Items.' *Methodology* 11 (1): 23–34.

Extensions and minor corrections since version 2021*

- The abstract has been revised and extended
- The shortened scales for *'Global self-esteem'* (*sel_m_fs*) and for *'Negative global self-esteem'* (*seld_m_fs*) used in later panel waves have been added to the scale appendix below and to the baseline survey data file (TREE2_Data_Wave_0_v2) in the data release (TREE, 2023).
- The reported Tucker coefficients for invariance across survey modes and survey settings include minor corrections.
- The student score variables *cogself1_fs*, *cogself2_fs*, *extregm_fs* have been renamed according to TREE naming conventions (to *cogselfa_fs*, *cogselfb_fs*, and *extreg_fs*).
- Some inconsistently used labels for scales and survey topics have been harmonised across the TREE2 data release (TREE, 2023).

* Sacchi, Stefan, Krebs-Oesch, Dominique (2021). Scaling methodology and scale reporting in the TREE2 panel survey. Documentation of scales implemented in the baseline survey (2016). University of Bern: TREE. http://dx.doi.org/10.48350/152055.

SCALE APPENDIX

Scales administered in the baseline survey

(Scale names linked with first page of scale-specific reporting)

Survey topics

Survey topics (continued) Baseline survey (2016)

Scale (or composite)	*Variable Name*	*AES Module*	*Source*	*Page*
9) Self-perception				
Global self-esteem	[sel_fs]	*Background*	*Rosenberg, 1979 (translated)*	128
Global self-esteem (shortened)	[sel_m_fs]	*Background*	*Rosenberg, 1979 (translated)*	130
Positive global self-esteem	[sele_fs]	*Background*	*Rosenberg, 1979 (translated)*	132
Negative global self-esteem	[seld_fs]	*Background*	*Rosenberg, 1979 (translated)*	134
Negative global self-esteem (shortened)	[seld_m_fs]	*Background*	*Rosenberg, 1979 (translated)*	136
General perceived self-efficacy scale (GSES)	[seef_fs]	*Background*	*GSES (adapted TREE1)*	138
Academic self-efficacy	[acaself_fs]	*General*	*Hascher, 2004*	140
Academic self-concept (PISA2000)	[scacad_fs]	*General*	*PISA 2000*	142
Verbal self-concept (PISA2000)	[scverb_fs]	*General*	*PISA 2000*	144
Maths self-concept [PISA 2000]	[matcon_fs]	*General*	*PISA 2000*	146
ICT self-concept	[ictabil_fs]	*Math*	*ICILS 2013*	148
Specific self-efficacy: numeracy	[selfeffa_fs]	*General [Math]*	*PISA 2012; Girnat, 2018*	150
Specific self-efficacy: algebra	[selfeffb_fs]	*General [Math]*	*PISA 2012; Girnat, 2018*	152
Specific self-efficacy: geometry	[selfeffc_fs]	*General [Math]*	*Girnat, 2018*	154
Specific self-efficacy: probability	[selfeffd_fs]	*General [Math]*	*Girnat, 2018*	156
10) Emotions related to maths classes				
Mathematics anxiety	[anxmath_fs]	*Math*	*PISA 2012*	158
Mathematics boredom	[boredom_fs]	*Math*	*AEQ-M (short-version)*	160
Mathematics anger	[anger_fs]	*Math*	*AEQ-M (short-version)*	162
Mathematics enjoyment	[enjoymath_fs]	*Math*	*AEQ-M (short-version)*	164
11) Volitional strategies				
Perseverance	[persev_fs]	*General*	*PISA 2012*	166
Effort: learning (PISA2000)	[effper_comp]	*Background*	*PISA2000*	168

Survey topics			Baseline survey (2016)	
Scale (or composite)	*Variable Name*	*AES Module*	*Source*	*Page*
12) Personality characteristics				
Big Five: extraversion	[big5_e_comp]	*Background*	*Rammstedt et al., 2014*	169
Big Five: agreeableness	[big5_a_comp]	*Background*	*Rammstedt et al., 2014*	169
Big Five: conscientiousness	[big5_c_comp]	*Background*	*Rammstedt et al., 2014*	169
Big Five: neuroticism	[big5_n_comp]	*Background*	*Rammstedt et al., 2014*	169
Big Five: openness	[big5_o_comp]	*Background*	*Rammstedt et al., 2014*	169
Internal locus of control	[loci_comp]	*Background*	*GESIS (short-version)*	170
External locus of control	[loce_comp]	*Background*	*GESIS (short-version)*	170
13) Values & attitudes				
Work-related extrinsic values	[vawe_fs]	*Background*	*TREE1 - based on Watermann, 2000*	172
Work-related intrinsic values	[vawi_fs]	*Background*	*TREE1 - based on Watermann, 2000*	174
Family values	[vafa_comp]	*Background*	*TREE1*	176
Positive attitude towards life	[posl_fs]	*AES Extension Survey*	*TREE1; Grob et al., 1991*	178
14) Attitudes related to mathematics classes				
Reality-based learning	[realref_fs]	*Math*	*Girnat, 2015, 2017*	180
Discovery / exploratory learning	[disclearn_fs]	*Math*	*Girnat, 2015, 2018*	182
Social learning	[soccomlearn_fs]	*Math*	*Girnat, 2015, 2019*	184
Social learning: social arrangement	[soclearn_fs]	*Math*	*Girnat, 2015, 2020*	186
Social learning: communication	[comlearn_fs]	*Math*	*Girnat, 2015, 2021*	188
Instructivist learning	[instreplearn_fs]	*Math*	*Girnat, 2015, 2022*	190
Instructivist learning: teachers instructions	[instrlearn_fs]	*Math*	*Girnat, 2015, 2023*	192
Instructivist learning: repetitive practice	[replearn_fs]	*Math*	*Girnat, 2015, 2024*	194
System aspect	[sysformasp_fs]	*Math*	*Girnat, 2015, 2025*	196
System aspect: logical thinking	[systasp_fs]	*Math*	*Girnat, 2015, 2026*	198
System aspect: formalism	[formasp_fs]	*Math*	*Girnat, 2015, 2027*	200
Scheme aspect	[schemasp_fs]	*Math*	*Girnat, 2015, 2028*	202
Application aspect	[applyasp_fs]	*Math*	*Girnat, 2015, 2029*	204

Survey topics

Emotional closeness to parents

Baseline survey sample

Composite descriptives Variable name	Mean	Std. dev.	Min.	Max.	Obs.
closep_comp	4.2	0.8	1	5	15664
Share of cases with imputed missing values:	3.5%				

35

Item descriptives Indicators	Mean	Std. dev.	Min.	Max.	Valid obs.
closef	4.1	1.1	1	5	15223
closem	4.4	0.9	1	5	15558

List of scales (wave 0)

Scale: Parental pressure to achieve

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	462	2	.000
	Baseline vs. saturated	20063	6	.000

2)	**Root mean squared error (RMSEA)**		.122
	90% Confidence interval:	lower bound	.113
	90% Confidence interval:	upper bound	.131
	Probability RMSEA <= 0.05		.000
3)	**Akaike's Information Criterion (AIC)**		142462
	Bayesian Information Criterion (BIC)		142554
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.977
	Tucker-Lewis Index (TLI)		.931
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.026
	Coefficient of determination (CD)		.816

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.811
(Cronbach's alpha = .751)	
McDonald's Omega	.811

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.95
factor 2	-.04
factor 3	-.09
factor 4	-.18

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
press1	0.69	0.01	0.68	0.70
press2	0.69	0.01	0.68	0.71
press3	0.78	0.00	0.77	0.79
press4	0.71	0.01	0.70	0.72

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
press1	2.2	1.0	1	4	15488
press2	3.0	0.9	1	4	15491
press3	3.0	0.8	1	4	15488
press4	2.8	0.9	1	4	15490

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
press1	1.66	-1.38	0.68	2.99
press2	1.79	-3.56	-1.79	0.80
press3	2.35	-5.01	-2.26	1.38
press4	1.84	-3.48	-1.23	1.53

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	1717	28	.000	105	14	.000	26	14	.027

Tests of measurement invariance across ...	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	31	6	.000	33	3	.000	11	3	.013
Strong invariance (plus equal intercepts)	923	6	.000	11	3	.010	4	3	.317
Strict invariance (plus equal error variances)	73	6	.000	12	3	.008	3	3	.413

Configural factor similarity across ...	**Survey languages**		**Survey settings**		**Survey modes**	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.999	classroom vs. unproctored	.998	web vs. PAP	.996
	French vs. Italian	.997				
	Italian vs. German	.993				

Factor score equivalence: group specific vs. invariant models for ...	**Survey languages**		**Survey settings**		**Survey modes**	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	1.000	unproctored	.999	PAP	.990
	Italian	.980				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
press_fs	0.0	0.9	-2.4	1.7	15535

Share of cases with imputed missing values: 0.6%

(Equivalence of scores from robust MLMV: CD = .997)

(Equivalence of Scores from Two-Step-Approach: CD = .984)

List of scales (wave 0)

Scale: Parents' achievement expectations

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	8040	2	.000
	Baseline vs. saturated	24621	6	.000
2)	**Root mean squared error (RMSEA)**			.606
	90% Confidence interval: lower bound			.595
	90% Confidence interval: upper bound			.617
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			77644
	Bayesian Information Criterion (BIC)			77731
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.673
	Tucker–Lewis Index (TLI)			.020
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.108
	Coefficient of determination (CD)			.854

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.837
(Cronbach's alpha = .774)	
McDonald's Omega	.834

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.35
Factor 2	.43
Factor 3	.11
Factor 4	-.19

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
expectf2	0.70	.007	0.69	0.72
expectf3	0.85	.005	0.84	0.86
expectm2	0.63	.009	0.62	0.65
expectm3	0.79	.005	0.78	0.80

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
expectf2	3.4	0.7	1	4	10568
expectf3	3.3	0.7	1	4	10566
expectm2	3.4	0.7	1	4	10862
expectm3	3.4	0.7	1	4	10864

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
expectf2	2.12	-5.87	-4.04	-0.32
expectf3	2.31	-5.88	-3.69	0.30
expectm2	1.75	-5.42	-3.28	0.14
expectm3	2.11	-6.40	-4.13	-0.12

List of scales (wave 0)

Scale: Parents' achievement expectations (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	297	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	15	6	.017
Strong invariance (plus equal intercepts)	126	6	.000
Strict invariance (plus equal error variances)	12	6	.072

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.996
Italian vs. German language version	.995

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.964

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
expectp_fs	0.0	0.9	-3.1	1.1	10952

Share of cases with imputed missing values: 4.3%

(Equivalence of scores from robust MLMV: CD = .991)

(Equivalence of scores from two-step approach: CD = .941)

List of scales (wave 0)

Scale: Mother's achievement expectations

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	4828	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			67851
	Bayesian Information Criterion (BIC)			67917
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.729

Reliability and Dimensionality

Ordinal Cronbach's Alpha .642

(Cronbach's alpha = .552)

McDonald's Omega .663

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.01
Factor 2	-.07
Factor 3	-.22

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
expectm1	0.42	.010	0.40	0.44
expectm2	0.80	.013	0.77	0.82
expectm3	0.65	.011	0.63	0.67

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
expectm1	2.8	0.8	1	4	10859
expectm2	3.4	0.7	1	4	10862
expectm3	3.4	0.7	1	4	10864

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
expectm1	0.83	-2.97	-0.79	1.48
expectm2	2.27	-6.07	-3.61	0.24
expectm3	1.68	-5.59	-3.50	-0.04

List of scales (wave 0)

Scale: Mother's achievement expectations (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	536	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	112	4	.000
Strong invariance (plus equal intercepts)	126	4	.000
Strict invariance (plus equal error variances)	66	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.965
French vs. Italian language version	.982
Italian vs. German language version	.979

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.964
Language: French	.961
Language: Italian	.970

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
expectm_fs	0.0	0.8	-2.8	1.2	10864

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .987)

(Equivalence of scores from two-step approach: CD = .957)

List of scales (wave 0)

Scale: Father's achievement expectations

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	7517	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			65854
	Bayesian Information Criterion (BIC)			65920
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.791

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.738
(Cronbach's alpha = .653)	
McDonald's Omega	.749

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.31
Factor 2	-.09
Factor 3	-.19

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
expectf1	0.55	.008	0.53	0.56
expectf2	0.83	.008	0.82	0.85
expectf3	0.72	.008	0.70	0.74

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
expectf1	2.9	0.9	1	4	10565
expectf2	3.4	0.7	1	4	10568
expectf3	3.3	0.7	1	4	10566

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
expectf1	1.17	-3.07	-1.05	1.32
expectf2	3.04	-7.28	-4.84	-0.32
expectf3	1.92	-5.13	-3.06	0.33

List of scales (wave 0)

Scale: Father's achievement expectations (continued) Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	429	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	100	4	.000
Strong invariance (plus equal intercepts)	57	4	.000
Strict invariance (plus equal error variances)	84	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.986
French vs. Italian language version	.997
Italian vs. German language version	.990

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.997
Language: French	.998
Language: Italian	.982

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
expectf_fs	0.0	0.8	-2.7	1.2	10569

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .988)

(Equivalence of scores from two-step approach: CD = .957)

List of scales (wave 0)

Scale: Mother's social norms about mathematics

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	12780	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			66659
	Bayesian Information Criterion (BIC)			66724
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.881

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.789
(Cronbach's alpha = .715)	
McDonald's Omega	.812

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.66
Factor 2	-.05
Factor 3	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
socnormsm1	0.87	.006	0.86	0.88
socnormsm2	0.89	.006	0.88	0.91
socnormsm3	0.50	.008	0.49	0.52

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
socnormsm1	3.2	0.7	1	4	10833
socnormsm2	3.1	0.8	1	4	10834
socnormsm3	2.4	0.9	1	4	10795

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
socnormsm1	3.95	-8.08	-4.66	1.62
socnormsm2	3.36	-5.95	-2.64	1.65
socnormsm3	0.99	-1.65	0.37	2.19

List of scales (wave 0)

Scale: Mother's social norms about mathematics (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	195	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	11	4	.030
Strong invariance (plus equal intercepts)	44	4	.000
Strict invariance (plus equal error variances)	80	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.998
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	.999
Language: French	.990
Language: Italian	.999

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
socnormsm_fs	0.1	0.9	-2.3	1.4	10847

Share of cases with imputed missing values: 0.6%

(Equivalence of scores from robust MLMV: CD = .996)

(Equivalence of scores from two-step approach: CD = .971)

List of scales (wave 0)

Scale: Father's social norms about mathematics

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	15486	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			60431
	Bayesian Information Criterion (BIC)			60496
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.922

Reliability and Dimensionality

Ordinal Cronbach's Alpha .837

(Cronbach's alpha = .771)

McDonald's Omega .851

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.85
Factor 2	-.04
Factor 3	-.14

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
socnormsf1	0.95	.004	0.94	0.96
socnormsf2	0.85	.005	0.84	0.86
socnormsf3	0.60	.007	0.59	0.62

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
socnormsf1	3.3	0.7	1	4	10576
socnormsf2	3.2	0.8	1	4	10572
socnormsf3	3.1	0.9	1	4	10567

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
socnormsf1	4.84	-9.33	-5.83	1.21
socnormsf2	3.14	-5.97	-3.09	1.20
socnormsf3	1.25	-2.99	-1.28	0.85

List of scales (wave 0)

Scale: Father's social norms about mathematics (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	198	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	15	4	.005
Strong invariance (plus equal intercepts)	85	4	.000
Strict invariance (plus equal error variances)	72	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.999
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	.999
Language: French	.996
Language: Italian	.956

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
socnormsf_fs	0.1	0.9	-2.4	1.2	10587

Share of cases with imputed missing values: 0.4%

(Equivalence of scores from robust MLMV: CD = .992)

(Equivalence of scores from two-step approach: CD = .960)

List of scales (wave 0)

Scale: Family educational support (PISA2000) — Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	16654	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			147278
	Bayesian Information Criterion (BIC)			147347
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.861

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.785
(Cronbach's alpha = .746)	
McDonald's Omega	.803

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.60
factor 2	-.07
factor 3	-.16

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
famedsup1	0.88	0.01	0.87	0.89
famedsup2	0.85	0.01	0.84	0.86
famedsup3	0.53	0.01	0.51	0.54

*** Note:** Replication of 'Famedsup'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
famedsup1	2.8	1.4	1	5	15462
famedsup2	2.6	1.4	1	5	15131
famedsup3	2.3	1.4	1	5	13709

List of scales (wave 0)

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	365	18	.000	101	9	.000	34	9	.000

Tests of measurement invariance across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	20	4	.001	9	2	.013	11	2	.005
Strong invariance (plus equal intercepts)	300	4	.000	32	2	.000	11	2	.003
Strict invariance (plus equal error variances)	12	4	.015	18	2	.000	2	2	.324

Configural factor similarity across …	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.998	classroom vs. unproctored	.999	web vs. PAP	.994
	French vs. Italian	.999				
	Italian vs. German	.999				

Factor score equivalence: group specific vs. invariant models for …	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	.999	unproctored	.998	PAP	.996
	Italian	.997				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
famedsup_fs	0.0	1.1	-1.6	2.2	15592

Share of cases with imputed missing values: 14.6%

(Equivalence of scores from robust MLMV: CD = .998)

List of scales (wave 0)

Scale: Social communication (PISA2000)

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	9734	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			124277
	Bayesian Information Criterion (BIC)			124346
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.750

Reliability and Dimensionality

Ordinal Cronbach's Alpha .723

(Cronbach's alpha = .647)

McDonald's Omega .729

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.24
factor 2	-.11
factor 3	-.20

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
soccom1	0.57	0.01	0.56	0.58
soccom2	0.71	0.01	0.69	0.72
soccom3	0.78	0.01	0.76	0.79

*** Note:** Replication of 'Soccom'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
soccom1	3.9	1.1	1	5	15566
soccom2	4.6	0.9	1	5	15570
soccom3	4.0	1.1	1	5	15555

List of scales (wave 0)

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	626	18	.000	611	9	.000	20	9	.017

Tests of measurement invariance across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	31	4	.000	26	2	.000	9	2	.012
Strong invariance (plus equal intercepts)	228	4	.000	107	2	.000	3	2	.231
Strict invariance (plus equal error variances)	92	4	.000	201	2	.000	3	2	.258

Configural factor similarity across …	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	1.000	classroom vs. unproctored	.995	web vs. PAP	.986
	French vs. Italian	.992				
	Italian vs. German	.988				

Factor score equivalence: group specific vs. invariant models for …	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	.998	web	.999
	French	1.000	unproctored	.980	PAP	.925
	Italian	.973				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
soccom_fs	0.0	0.5	-2.1	0.5	15588

Share of cases with imputed missing values: 0.4%

(Equivalence of scores from robust MLMV: CD = .986)

List of scales (wave 0)

Scale: Social communication (adapted TREE2)

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	26651	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			119342
	Bayesian Information Criterion (BIC)			119411
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.890

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.889
(Cronbach's alpha = .851)	
McDonald's Omega	.889

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.06
factor 2	-.11
factor 3	-.11

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf.	interval]
soccom3 **	0.84	0.00	0.84	0.85
soccom4	0.86	0.00	0.85	0.86
soccom5	0.86	0.00	0.86	0.87

*** Note:** Scale from TREE1 / PISA2000 adapted for TREE2

**** Note:** Original item from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
soccom3 **	4.0	1.1	1	5	15555
soccom4	3.9	1.2	1	5	15560
soccom5	4.0	1.1	1	5	15563

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	942	18	.000	159	9	.000	49	9	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	50	4	.000	5	2	.094	2	2	.459
Strong invariance (plus equal intercepts)	129	4	.000	37	2	.000	2	2	.408
Strict invariance (plus equal error variances)	211	4	.000	19	2	.000	6	2	.041

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.999	classroom vs. unproctored	1.000	web vs. PAP	1.000
	French vs. Italian	.999				
	Italian vs. German	.997				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	1.000	unproctored	1.000	PAP	1.000
	Italian	.997				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
soccom_m_fs	0.0	0.9	-2.6	0.9	15591

Share of cases with imputed missing values: 0.5%

(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Perceived social network support

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	2147	5	.000
	Baseline vs. saturated	58182	10	.000

2)	**Root mean squared error (RMSEA)**	.169
	90% Confidence interval: lower bound	.163
	90% Confidence interval: upper bound	.175
	Probability RMSEA <= 0.05	.000
3)	**Akaike's Information Criterion (AIC)**	233311
	Bayesian Information Criterion (BIC)	233425
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.963
	Tucker-Lewis Index (TLI)	.926
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.035
	Coefficient of determination (CD)	.939

Reliability and Dimensionality

Ordinal Cronbach's Alpha .920

(Cronbach's alpha = .896)

McDonald's Omega .920

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	3.45
factor 2	.09
factor 3	.00
factor 4	-.06
factor 5	-.12

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
closupp1	0.81	0.00	0.80	0.81
closupp2	0.93	0.00	0.93	0.93
closupp3	0.88	0.00	0.88	0.88
closupp4	0.68	0.00	0.67	0.69
closupp5	0.86	0.00	0.86	0.87

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
closupp1	5.4	1.6	1	7	14695
closupp2	5.6	1.6	1	7	14756
closupp3	5.7	1.6	1	7	14760
closupp4	5.1	1.7	1	7	14086
closupp5	5.5	1.8	1	7	14430

List of scales (wave 0)

Scale: Perceived social network support (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	635	40	.000	802	20	.000	105	20	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	33	8	.000	87	4	.000	8	4	.075
Strong invariance (plus equal intercepts)	205	8	.000	219	4	.000	13	4	.014
Strict invariance (plus equal error variances)	291	8	.000	17	4	.002	26	4	.000

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	1.000	classroom vs. unproctored	.998	web vs. PAP	.999
	French vs. Italian	.999				
	Italian vs. German	.999				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	1.000	unproctored	.999	PAP	1.000
	Italian	1.000				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
closupp_fs	0.0	1.2	-3.9	1.2	15034

Share of cases with imputed missing values: 10.4%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Parental reading interest

Baseline survey sample

Composite descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
joyreadp_comp	3.1	0.8	1	4	15244

Share of cases with imputed missing values: 8.7%
(Including "don't know"-answers for one parent)

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid obs.
joyreadm	3.4	0.9	1	4	15004
joyreadf	2.9	1.1	1	4	14164

List of scales (wave 0)

Scale: Cultural communication (PISA2000)

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	8034	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			146251
	Bayesian Information Criterion (BIC)			146320
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.727

Reliability and Dimensionality

Ordinal Cronbach's Alpha .677

(Cronbach's alpha = .606)

McDonald's Omega .690

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.11
factor 2	-.10
factor 3	-.21

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf.	interval]
cultcom1	0.72	0.01	0.70	0.73
cultcom2	0.75	0.01	0.74	0.77
cultcom3	0.47	0.01	0.45	0.49

*** Note:** Replication of 'Cultcom'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
cultcom1	3.0	1.3	1	5	15593
cultcom2	3.2	1.3	1	5	15578
cultcom3	1.7	1.2	1	5	15575

List of scales (wave 0)

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	369	18	.000	267	9	.000	42	9	.000

Tests of measurement invariance across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	16	4	.003	8	2	.019	1	2	.673
Strong invariance (plus equal intercepts)	263	4	.000	141	2	.000	14	2	.001
Strict invariance (plus equal error variances)	30	4	.000	15	2	.001	13	2	.002

Configural factor similarity across …	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.998	classroom vs. unproctored	.999	web vs. PAP	.998
	French vs. Italian	.987				
	Italian vs. German	.996				

Factor score equivalence: group specific vs. invariant models for …	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	.999	classroom	1.000	web	1.000
	French	.996	unproctored	.999	PAP	.996
	Italian	.970				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cultcom_fs	0.0	0.8	-1.6	1.8	15601

Share of cases with imputed missing values: 0.3%

(Equivalence of scores from robust MLMV: CD = .998)

List of scales (wave 0)

Scale: Cultural communication (adapted TREE2)

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	16199	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			137695
	Bayesian Information Criterion (BIC)			137764
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.829

Reliability and Dimensionality

Ordinal Cronbach's Alpha .811
(Cronbach's alpha = .762)
McDonald's Omega .814

Test of (one-)dimensionality (parallel analysis)
Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.63
factor 2	-.11
factor 3	-.17

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf.	interval]
cultcom1 **	0.80	0.00	0.79	0.81
cultcom2 **	0.68	0.01	0.67	0.69
cultcom4	0.83	0.00	0.82	0.84

*** Note:** Scale from TREE1 / PISA2000 adapted for TREE2
**** Note:** Original items from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
cultcom1 **	3.0	1.3	1	5	15593
cultcom2 **	3.2	1.3	1	5	15578
cultcom4	3.8	1.1	1	5	15571

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	313	18	.000	333	9	.000	26	9	.002

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	14	4	.008	8	2	.015	5	2	.073
Strong invariance (plus equal intercepts)	206	4	.000	212	2	.000	1	2	.519
Strict invariance (plus equal error variances)	30	4	.000	24	2	.000	7	2	.032

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	1.000	classroom vs. unproctored	.999	web vs. PAP	.997
	French vs. Italian	.997				
	Italian vs. German	.996				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	1.000	unproctored	.999	PAP	.998
	Italian	.996				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cultcom_m_fs	0.0	0.9	-2.4	1.5	15610

Share of cases with imputed missing values: 0.4%
(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Household possessions: classical culture (PISA2000) — Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	11545	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			52733
	Bayesian Information Criterion (BIC)			52802
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.817

Reliability and Dimensionality

Ordinal Cronbach's Alpha .720
(Cronbach's alpha = .556)
McDonald's Omega .742

Test of (one-)dimensionality (parallel analysis)
Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.30
factor 2	-.06
factor 3	-.20

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
cultposs1	0.74	0.01	0.73	0.76
cultposs2	0.86	0.01	0.85	0.88
cultposs3	0.46	0.01	0.45	0.48

*** Note:** Replication of 'Cultposs'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
cultposs1	0.4	0.5		1	15977
cultposs2	0.4	0.5		1	15990
cultposs3	0.7	0.4		1	16009

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
cultposs1	1.90	0.71		
cultposs2	3.51	0.55		
cultposs3	0.91	-1.23		

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	4574	18	.000	101	9	.000	79	9	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	53	4	.000	1	2	.759	13	2	.002
Strong invariance (plus equal intercepts)	887	4	.000	52	2	.000	21	2	.000
Strict invariance (plus equal error variances)	366	4	.000	21	2	.000	19	2	.000

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.996	classroom vs. unproctored	1.000	web vs. PAP	.996
	French vs. Italian	1.000				
	Italian vs. German	.997				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	.979	classroom	1.000	web	.999
	French	.890	unproctored	1.000	PAP	.985
	Italian	.819				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cultposs_fs	0.0	0.8	-1.0	1.1	16028

Share of cases with imputed missing values: 0.5%
(Equivalence of scores from robust MLMV: CD = .969)
(Equivalence of Scores from Two-Step-Approach: CD = .96)

*** Note:** The calculation of model-based invariance tests requires that we constrain the error variance of *cultposs2* to zero.

List of scales (wave 0)

Scale: Embodied cultural capital

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	1455	9	.000
	Baseline vs. saturated	42913	15	.000
2)	**Root mean squared error (RMSEA)**			.101
	90% Confidence interval: lower bound			.096
	90% Confidence interval: upper bound			.105
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			166162
	Bayesian Information Criterion (BIC)			166300
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.966
	Tucker-Lewis Index (TLI)			.944
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.033
	Coefficient of determination (CD)			.883

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.870
(Cronbach's alpha = .822)	
McDonald's Omega	.872

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	3.13
factor 2	.11
factor 3	-.04
factor 4	-.05
factor 5	-.12
factor 6	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
manners1	0.53	0.01	0.52	0.55
manners2	0.80	0.00	0.80	0.81
manners3	0.74	0.00	0.73	0.75
verbskill1	0.75	0.00	0.74	0.76
verbskill2	0.78	0.00	0.78	0.79
verbskill3	0.75	0.00	0.74	0.75

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
manners1	3.0	0.8	1	4	15819
manners2	3.1	0.7	1	4	15805
manners3	3.1	0.7	1	4	15807
verbskill1	3.0	0.7	1	4	15827
verbskill2	3.0	0.8	1	4	15817
verbskill3	2.9	0.7	1	4	15776

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
manners1	1.21	-3.68	-1.95	1.19
manners2	2.57	-6.65	-2.90	1.98
manners3	2.10	-6.12	-2.90	1.50
verbskill1	2.13	-5.28	-2.04	1.80
verbskill2	2.39	-5.71	-2.08	1.73
verbskill3	2.13	-5.33	-1.79	2.15

List of scales (wave 0)

Scale: Embodied cultural capital (continued) — Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	765	54	.000	221	27	.000	63	27	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	21	10	.018	36	5	.000	14	5	.018
Strong invariance (plus equal intercepts)	70	10	.000	24	5	.000	10	5	.085
Strict invariance (plus equal error variances)	197	10	.000	57	5	.000	15	5	.011

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	1.000	classroom vs. unproctored	.999	web vs. PAP	.997
	French vs. Italian	.999				
	Italian vs. German	.999				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	1.000	unproctored	.999	PAP	.998
	Italian	.999				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
inccap_fs	0.0	0.9	-3.2	1.8	15846

Share of cases with imputed missing values: 0.9%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of Scores from Two-Step-Approach: CD = .989)

List of scales (wave 0)

Scale: Embodied cultural capital: manners

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	12618	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			88215
	Bayesian Information Criterion (BIC)			88284
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.798

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.763
(Cronbach's alpha = .684)	
McDonald's Omega	.769

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.41
factor 2	-.10
factor 3	-.20

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
manners1	0.60	0.01	0.58	0.61
manners2	0.74	0.01	0.73	0.76
manners3	0.83	0.01	0.81	0.84

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
manners1	3.0	0.8	1	4	15819
manners2	3.1	0.7	1	4	15805
manners3	3.1	0.7	1	4	15807

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
manners1	1.41	-3.87	-2.07	1.28
manners2	2.10	-5.87	-2.59	1.77
manners3	2.85	-7.40	-3.62	1.88

List of scales (wave 0)

Scale: Embodied cultural capital: manners (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	470	18	.000	138	9	.000	15	9	.082

Tests of measurement invariance across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	7	4	.160	1	2	.751	3	2	.231
Strong invariance (plus equal intercepts)	28	4	.000	16	2	.000	3	2	.280
Strict invariance (plus equal error variances)	40	4	.000	14	2	.001	4	2	.119

Configural factor similarity across …	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.999	classroom vs. unproctored	1.000	web vs. PAP	.998
	French vs. Italian	.999				
	Italian vs. German	.999				

Factor score equivalence: group specific vs. invariant models for …	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	.998	unproctored	1.000	PAP	.998
	Italian	.997				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
manners_fs	0.0	0.8	-2.8	1.5	15843

Share of cases with imputed missing values: 0.5%

(Equivalence of scores from robust MLMV: CD = .998)

(Equivalence of Scores from Two-Step-Approach: CD = .988)

List of scales (wave 0)

Scale: Embodied cultural capital: verbal skills

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	16621	3	.000

2)	**Root mean squared error (RMSEA)**		.000
	90% Confidence interval:	lower bound	.000
	90% Confidence interval:	upper bound	.000
	Probability RMSEA <= 0.05		1.000
3)	**Akaike's Information Criterion (AIC)**		90127
	Bayesian Information Criterion (BIC)		90196
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		1.000
	Tucker-Lewis Index (TLI)		1.000
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.000
	Coefficient of determination (CD)		.821

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.818
(Cronbach's alpha = .759)	
McDonald's Omega	.819

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.64
factor 2	-.14
factor 3	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
verbskill1	0.74	0.00	0.73	0.75
verbskill2	0.80	0.00	0.79	0.81
verbskill3	0.79	0.00	0.78	0.80

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
verbskill1	3.0	0.7	1	4	15827
verbskill2	3.0	0.8	1	4	15817
verbskill3	2.9	0.7	1	4	15776

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
verbskill1	2.03	-5.16	-2.00	1.78
verbskill2	2.49	-5.91	-2.15	1.82
verbskill3	2.43	-5.80	-1.96	2.36

List of scales (wave 0)

Scale: Embodied cultural capital: verbal skills (continued) — Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	209	18	.000	24	9	.005	34	9	.000

Tests of measurement invariance across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	6	4	.227	4	2	.137	12	2	.003
Strong invariance (plus equal intercepts)	36	4	.000	2	2	.425	4	2	.106
Strict invariance (plus equal error variances)	89	4	.000	13	2	.002	8	2	.023

Configural factor similarity across ...

Tucker's congruence coefficient

Survey languages	TCC	Survey settings	TCC	Survey modes	TCC
German vs. French	1.000	classroom vs. unproctored	1.000	web vs. PAP	.995
French vs. Italian	.998				
Italian vs. German	.999				

Factor score equivalence: group specific vs. invariant models for ...

Coefficient of determination

Survey languages	CD	Survey settings	CD	Survey modes	CD
German	1.000	classroom	1.000	web	1.000
French	1.000	unproctored	1.000	PAP	.993
Italian	.998				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
verbskill_fs	0.0	0.9	-2.7	1.6	15841

Share of cases with imputed missing values: 0.6%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of Scores from Two-Step-Approach: CD = .992)

List of scales (wave 0)

Scale: Cultural activities

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	7949	14	.000
	Baseline vs. saturated	27943	21	.000

2)	**Root mean squared error (RMSEA)**	.189
	90% Confidence interval: lower bound	.186
	90% Confidence interval: upper bound	.193
	Probability RMSEA <= 0.05	.000
3)	**Akaike's Information Criterion (AIC)**	260288
	Bayesian Information Criterion (BIC)	260449
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.716
	Tucker-Lewis Index (TLI)	.574
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.118
	Coefficient of determination (CD)	.809

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.743
(Cronbach's alpha = .668)	
McDonald's Omega	.726

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.14
factor 2	.76
factor 3	.03
factor 4	-.02
factor 5	-.13
factor 6	-.20
factor 7	-.20

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
cult1 **	0.36	0.01	0.34	0.37
cult2 **	0.70	0.01	0.69	0.71
cult3 **	0.50	0.01	0.48	0.51
cult4 **	0.77	0.00	0.76	0.78
cult5 **	0.74	0.01	0.73	0.75
cult7	0.29	0.01	0.27	0.31
cult9	0.24	0.01	0.23	0.26

*** Note:** Scale from TREE1 / PISA2000 adapted for TREE2

**** Note:** Original items from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
cult1 **	2.8	1.0	1	4	15787
cult2 **	1.8	0.9	1	4	15776
cult3 **	1.6	0.8	1	4	15769
cult4 **	1.3	0.6	1	4	15771
cult5 **	1.6	0.7	1	4	15761
cult7	2.6	1.0	1	4	15766
cult9	2.4	1.2	1	4	15761

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
cult1	0.83	-2.69	-0.45	0.93
cult2	1.54	-0.32	2.13	3.59
cult3	1.17	0.43	2.48	3.64
cult4	1.93	2.19	4.18	5.39
cult5	1.76	0.12	3.13	4.74
cult7	0.70	-1.83	0.18	1.41
cult9	0.60	-0.93	0.30	1.17

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	1553	70	.000	737	35	.000	149	35	.000

Tests of measurement invariance across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	107	12	.000	30	6	.000	19	6	.005
Strong invariance (plus equal intercepts)	1198	12	.000	231	6	.000	74	6	.000
Strict invariance (plus equal error variances)	142	12	.000	269	6	.000	35	6	.000

Configural factor similarity across ...

Tucker's congruence coefficient

Survey languages	TCC	Survey settings	TCC	Survey modes	TCC
German vs. French	.992	classroom vs. unproctored	.997	web vs. PAP	.987
French vs. Italian	.996				
Italian vs. German	.992				

Factor score equivalence: group specific vs. invariant models for ...

Coefficient of determination

Survey languages	CD	Survey settings	CD	Survey modes	CD
German	1.000	classroom	1.000	web	1.000
French	.995	unproctored	.997	PAP	.990
Italian	.998				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cult_fs	0.0	0.8	-1.8	3.1	15797

Share of cases with imputed missing values: 0.6%

(Equivalence of scores from robust MLMV: CD = .977)

(Equivalence of Scores from Two-Step-Approach: CD = .886)

List of scales (wave 0)

Scale: Lowbrow cultural activities

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	7348	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			124416
	Bayesian Information Criterion (BIC)			124485
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.728

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.668
(Cronbach's alpha = .599)	
McDonald's Omega	.679

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.05
factor 2	-.10
factor 3	-.22

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
cult3 *	0.54	0.01	0.52	0.56
cult7	0.58	0.01	0.56	0.59
cult9	0.80	0.01	0.78	0.82

*** Note:** Original item from TREE1 / PISA2000

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
cult3 *	1.6	0.8	1	4	15769
cult7	2.6	1.0	1	4	15766
cult9	2.4	1.2	1	4	15761

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
cult3	1.11	0.43	2.46	3.56
cult7	1.27	-2.14	0.17	1.64
cult9	2.25	-1.53	0.47	1.88

List of scales (wave 0)

Scale: Lowbrow cultural activities (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	993	18	.000	164	9	.000	50	9	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	65	4	.000	18	2	.000	13	2	.002
Strong invariance (plus equal intercepts)	674	4	.000	107	2	.000	24	2	.000
Strict invariance (plus equal error variances)	162	4	.000	13	2	.002	5	2	.071

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.985	classroom vs. unproctored	.996	web vs. PAP	.949
	French vs. Italian	.999				
	Italian vs. German	.989				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	.992	classroom	.999	web	.999
	French	.975	unproctored	.990	PAP	.852
	Italian	.996				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cultlow_fs	0.0	0.8	-1.4	1.8	15788

Share of cases with imputed missing values: 0.3%

(Equivalence of scores from robust MLMV: CD = .99)

(Equivalence of Scores from Two-Step-Approach: CD = .975)

List of scales (wave 0)

Scale: Highbrow cultural activities [PISA 2000]

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	14402	3	.000

2)	**Root mean squared error (RMSEA)**	.000
	90% Confidence interval: lower bound	.000
	90% Confidence interval: upper bound	.000
	Probability RMSEA <= 0.05	1.000
3)	**Akaike's Information Criterion (AIC)**	90498
	Bayesian Information Criterion (BIC)	90567
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	1.000
	Tucker-Lewis Index (TLI)	1.000
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.000
	Coefficient of determination (CD)	.805

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.793
(Cronbach's alpha = .690)	
McDonald's Omega	.795

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.53
factor 2	-.13
factor 3	-.17

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
cult2	0.69	0.01	0.68	0.70
cult4	0.82	0.01	0.81	0.83
cult5	0.74	0.01	0.73	0.75

*** Note:** Replication of 'Cultactv'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
cult2	1.8	0.9	1	4	15776
cult4	1.3	0.6	1	4	15771
cult5	1.6	0.7	1	4	15761

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
cult2	1.69	-0.33	2.26	3.75
cult4	2.53	2.64	4.95	6.28
cult5	2.01	0.15	3.41	5.05

List of scales (wave 0)

Scale: Highbrow cultural activities [PISA 2000] (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	283	18	.000	436	9	.000	58	9	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	40	4	.000	5	2	.085	1	2	.518
Strong invariance (plus equal intercepts)	125	4	.000	48	2	.000	10	2	.008
Strict invariance (plus equal error variances)	48	4	.000	176	2	.000	13	2	.001

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.997	classroom vs. unproctored	.999	web vs. PAP	.999
	French vs. Italian	.999				
	Italian vs. German	.999				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	.993	unproctored	.999	PAP	.996
	Italian	.999				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
culthigh_fs	0.0	0.8	-0.9	2.6	15788

Share of cases with imputed missing values: 0.3%

(Equivalence of scores from robust MLMV: CD = .98)

(Equivalence of Scores from Two-Step-Approach: CD = .886)

List of scales (wave 0)

Scale: Household Possessions: Family Wealth (PISA2000) — Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	12119	27	.000
	Baseline vs. saturated	41971	36	.000
2)	**Root mean squared error (RMSEA)**			.167
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			138697
	Bayesian Information Criterion (BIC)			138904
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.712
	Tucker-Lewis Index (TLI)			.616
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.079
	Coefficient of determination (CD)			.839

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.782
(Cronbach's alpha = .565)	
McDonald's Omega	.789

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.83
factor 2	.49
factor 3	.40
factor 4	.08
factor 5	.07
factor 6	.02
factor 7	-.10
factor 8	-.15
factor 9	-.25

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
wealth1	0.71	0.00	0.70	0.72
wealth2	0.57	0.01	0.56	0.58
wealth3	0.31	0.01	0.29	0.32
wealth4	0.81	0.00	0.80	0.82
wealthn1	0.59	0.01	0.58	0.61
wealthn2	0.35	0.01	0.33	0.36
wealthn3	0.50	0.01	0.49	0.51
wealthn4	0.42	0.01	0.41	0.44
wealthn5	0.55	0.01	0.54	0.56

* **Note:** Replication of 'Wealth'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
wealth1	0.9	0.3		1	16040
wealth2	0.9	0.3		1	16039
wealth3	0.6	0.5		1	15942
wealth4	1.0	0.1		1	16043
wealthn1	3.9	0.4	1	4	16037
wealthn2	2.8	0.8	1	4	16037
wealthn3	3.3	0.8	1	4	16032
wealthn4	2.7	0.8	1	4	16030
wealthn5	2.9	0.7	1	4	16037

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
wealth1	1.64	-3.46		
wealth2	1.08	-2.75		
wealth3	0.29	-0.36		
wealth4	1.76	-5.87		
wealthn1	1.46	-6.37	-4.51	-3.29
wealthn2	0.79	-3.58	-0.51	1.35
wealthn3	1.01	-4.94	-1.65	-0.01
wealthn4	1.18	-3.18	-0.25	2.19
wealthn5	1.48	-6.23	-1.26	2.00

List of scales (wave 0)

Scale: Household Possessions: Family Wealth (PISA2000) (cont.) Baseline survey sample

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	4879	108	.000	1025	54	.000	1065	54	.000

Tests of measurement invariance across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	139	16	.000	92	8	.000	103	8	.000
Strong invariance (plus equal intercepts)	499	16	.000	74	8	.000	44	8	.000
Strict invariance (plus equal error variances)	1367	16	.000	270	8	.000	147	8	.000

Configural factor similarity across …	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.989	classroom vs. unproctored	.997	web vs. PAP	.965
	French vs. Italian	.992				
	Italian vs. German	.991				

Factor score equivalence: group specific vs. invariant models for …	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	.997
	French	.999	unproctored	.995	PAP	.964
	Italian	.959				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
wealth_fs	0.0	0.8	-4.0	1.8	16057

Share of cases with imputed missing values: 1.0%

(Equivalence of scores from robust MLMV: CD = .641)

(Equivalence of Scores from Two-Step-Approach: CD = .508)

List of scales (wave 0)

Scale: Household Possessions: Family Wealth (adapted TREE2) — Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	8521	14	.000
	Baseline vs. saturated	38309	21	.000
2)	**Root mean squared error (RMSEA)**			.195
	90% Confidence interval: lower bound			.191
	90% Confidence interval: upper bound			.198
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			59604
	Bayesian Information Criterion (BIC)			59765
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.778
	Tucker-Lewis Index (TLI)			.667
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.079
	Coefficient of determination (CD)			.837

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.813
(Cronbach's alpha = .548)	
McDonald's Omega	.815

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.76
factor 2	.46
factor 3	.20
factor 4	.02
factor 5	-.07
factor 6	-.12
factor 7	-.24

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
wealth1 **	0.77	0.00	0.76	0.77
wealth2 **	0.62	0.01	0.61	0.63
wealth4 **	0.75	0.00	0.74	0.76
wealth5	0.61	0.01	0.60	0.62
wealthn1 **	0.51	0.01	0.50	0.52
wealthn3 **	0.47	0.01	0.46	0.49
wealthn5 **	0.60	0.01	0.59	0.61

*** Note:** Scale from TREE1 / PISA2000 adapted for TREE2

**** Note:** Original items from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
wealth1 **	0.9	0.3		1	16040
wealth2 **	0.9	0.3		1	16039
wealth4 **	1.0	0.1		1	16043
wealth5	0.7	0.5		1	16021
wealthn1 **	3.9	0.4	1	4	16037
wealthn3 **	3.3	0.8	1	4	16032
wealthn5 **	2.9	0.7	1	4	16037

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
wealth1	2.07	-3.91		
wealth2	1.43	-3.03		
wealth4	2.04	-6.28		
wealth5	1.44	-0.76		
wealthn1	1.07	-5.80	-4.09	-2.96
wealthn3	0.87	-4.81	-1.60	-0.01
wealthn5	1.79	-6.65	-1.40	2.20

List of scales (wave 0)

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	2014	70	.000	777	35	.000	890	35	.000

Tests of measurement invariance across …	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	168	12	.000	144	6	.000	74	6	.000
Strong invariance (plus equal intercepts)	329	12	.000	65	6	.000	25	6	.000
Strict invariance (plus equal error variances)	983	12	.000	175	6	.000	140	6	.000

Configural factor similarity across …	**Survey languages**		**Survey settings**		**Survey modes**	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.996	classroom vs. unproctored	.992	web vs. PAP	.965
	French vs. Italian	.975				
	Italian vs. German	.989				

Factor score equivalence: group specific vs. invariant models for …	**Survey languages**		**Survey settings**		**Survey modes**	
Coefficient of determination		CD		CD		CD
	German	.999	classroom	.999	web	.999
	French	.978	unproctored	.991	PAP	.947
	Italian	.902				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
wealth_m_fs	0.0	0.8	-3.6	1.3	16056

Share of cases with imputed missing values: 0.4%

(Equivalence of scores from robust MLMV: CD = .83)

(Equivalence of Scores from Two-Step-Approach: CD = .692)

List of scales (wave 0)

Family affluence scale (FASIII)

Baseline survey sample

Composite descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
fasiii_comp	9.5	2.1	0	13	16059

Share of cases with imputed missing values: 0.5%

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid obs.	
wealthn4	1.5	0.6	0	2	16030	*
wealth2	0.9	0.3	0	1	16039	
wealthn3	2.3	0.8	0	3	16032	*
wealthn5	1.9	0.7	0	3	16037	*
wealth1	0.9	0.3	0	1	16040	
holyn	1.9	1.0	0	3	16028	*

* Items recoded for composite calculation (see Hobza et al. 2017)

List of scales (wave 0)

Scale: Capabilities

Baseline survey sample

Model and Fit Statistics

1) **Likelihood-ratio tests**

	chi2	df	p > chi2
Model vs. saturated	1666	5	.000
Baseline vs. saturated	37134	10	.000

2) **Root mean squared error (RMSEA)** .145

90% Confidence interval:	lower bound	.139
90% Confidence interval:	upper bound	.151
Probability RMSEA <= 0.05		.000

3) **Akaike's Information Criterion (AIC)** 221347

Bayesian Information Criterion (BIC) 221462

4) **Baseline comparison**

Comparative Fit Index (CFI)	.955
Tucker-Lewis Index (TLI)	.911

5) **Size of residuals**

Stand. root mean squared residual (SRMR)	.038
Coefficient of determination (CD)	.874

Reliability and Dimensionality

Ordinal Cronbach's Alpha .871

(Cronbach's alpha = .845)

McDonald's Omega .871

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.79
factor 2	.10
factor 3	-.07
factor 4	-.13
factor 5	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
cap1	0.76	0.00	0.75	0.77
cap2	0.78	0.00	0.77	0.79
cap3	0.79	0.00	0.78	0.80
cap4	0.69	0.00	0.68	0.70
cap5	0.76	0.00	0.75	0.77

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
cap1	5.9	1.3	1	7	15756
cap2	5.7	1.2	1	7	15733
cap3	5.9	1.2	1	7	15732
cap4	5.3	1.3	1	7	15714
cap5	5.7	1.2	1	7	15738

List of scales (wave 0)

Tests and Indices of Factorial Invariance across …

Equality of the variance-covariance matrices across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	1233	40	.000	412	20	.000	32	20	.042

Tests of measurement invariance across …	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	106	8	.000	21	4	.000	7	4	.145
Strong invariance (plus equal intercepts)	601	8	.000	75	4	.000	11	4	.025
Strict invariance (plus equal error variances)	216	8	.000	15	4	.005	4	4	.456

Configural factor similarity across …	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.996	classroom vs. unproctored	.999	web vs. PAP	.998
	French vs. Italian	.997				
	Italian vs. German	.997				

Factor score equivalence: group specific vs. invariant models for …	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	.998	unproctored	1.000	PAP	.998
	Italian	.999				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cap_fs	0.0	0.9	-4.3	1.2	15783

Share of cases with imputed missing values: 0.7%

(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Positive attitude towards school

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	22788	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			205667
	Bayesian Information Criterion (BIC)			205739
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.835

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.809
(Cronbach's alpha = .784)	
McDonald's Omega	.813

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.61
Factor 2	-.10
Factor 3	-.17

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
posatt1	0.74	.004	0.73	0.75
posatt2	0.86	.004	0.85	0.87
posatt3	0.70	.004	0.69	0.71

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
posatt1	3.8	1.3	1	6	22295
posatt2	4.1	1.3	1	6	22288
posatt3	4.6	1.3	1	6	22287

List of scales (wave 0)

Scale: Positive attitude towards school (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	998	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	17	4	.002
Strong invariance (plus equal intercepts)	172	4	.000
Strict invariance (plus equal error variances)	217	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	1.000
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
posatt_fs	0.0	0.9	-2.5	1.4	22299

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Enjoyment in school

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	24844	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			216963
	Bayesian Information Criterion (BIC)			217035
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.856

Reliability and Dimensionality

Ordinal Cronbach's Alpha .821
(Cronbach's alpha = .796)
McDonald's Omega .825

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.67
Factor 2	-.08
Factor 3	-.16

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
enjoyschool1	0.76	.004	0.75	0.77
enjoyschool2	0.89	.004	0.88	0.89
enjoyschool3	0.69	.004	0.68	0.70

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
enjoyschool1	3.2	1.5	1	6	22254
enjoyschool2	3.5	1.4	1	6	22252
enjoyschool3	3.9	1.4	1	6	22257

List of scales (wave 0)

Scale: Enjoyment in school (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	506	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	33	4	.000
Strong invariance (plus equal intercepts)	258	4	.000
Strict invariance (plus equal error variances)	34	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.992
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.998
Language: Italian	.994

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
enjoyschool_fs	0.0	1.1	-2.1	2.1	22267

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Physical complaints in school

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	29	2	.000
	Baseline vs. saturated	36796	6	.000

2)	**Root mean squared error (RMSEA)**	.025
	90% Confidence interval: lower bound	.017
	90% Confidence interval: upper bound	.033
	Probability RMSEA <= 0.05	1.000
3)	**Akaike's Information Criterion (AIC)**	272002
	Bayesian Information Criterion (BIC)	272098
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.999
	Tucker–Lewis Index (TLI)	.998
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.005
	Coefficient of determination (CD)	.857

Reliability and Dimensionality

Ordinal Cronbach's Alpha .847
(Cronbach's alpha = .772)
McDonald's Omega .849

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.22
Factor 2	-.09
Factor 3	-.10
Factor 4	-.12

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
physpain1	0.78	.003	0.77	0.79
physpain2	0.79	.003	0.78	0.79
physpain3	0.82	.003	0.81	0.82
physpain4	0.67	.004	0.66	0.68

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
physpain1	1.7	1.3	1	6	22260
physpain2	1.7	1.4	1	6	22249
physpain3	1.7	1.3	1	6	22222
physpain4	2.3	1.6	1	6	22245

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1179	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	76	6	.000
Strong invariance (plus equal intercepts)	188	6	.000
Strict invariance (plus equal error variances)	542	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.997
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.988

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
physpain_fs	0.0	0.8	-.6	3.5	22271

Share of cases with imputed missing values: 0.3%
(Equivalence of scores from robust MLMV: CD = .995)

List of scales (wave 0)

Scale: Worries about school

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	21848	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			240309
	Bayesian Information Criterion (BIC)			240381
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.836

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.795
(Cronbach's alpha = .753)	
McDonald's Omega	.802

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.57
Factor 2	-.09
Factor 3	-.18

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
trouschool1	0.78	.004	0.78	0.79
trouschool2	0.86	.004	0.85	0.87
trouschool3	0.62	.005	0.61	0.63

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
trouschool1	2.9	1.6	1	6	22260
trouschool2	3.2	1.7	1	6	22263
trouschool3	3.4	1.9	1	6	22263

List of scales (wave 0)

Scale: Worries about school (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1522	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	51	4	.000
Strong invariance (plus equal intercepts)	889	4	.000
Strict invariance (plus equal error variances)	295	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.999
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.996

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
trouschool_fs	0.0	1.2	-1.9	2.5	22270

Share of cases with imputed missing values: 0.1%
(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Social problems in school

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	39687	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			164458
	Bayesian Information Criterion (BIC)			164530
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.929

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.886
(Cronbach's alpha = .817)	
McDonald's Omega	.889

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.07
Factor 2	-.05
Factor 3	-.12

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
socprob1	0.95	.002	0.95	0.95
socprob2	0.84	.003	0.84	0.85
socprob3	0.76	.003	0.75	0.77

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
socprob1	1.5	1.0	1	6	22244
socprob2	1.7	1.2	1	6	22259
socprob3	1.5	1.1	1	6	22239

List of scales (wave 0)

Scale: Social problems in school (continued) — Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	466	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	16	4	.003
Strong invariance (plus equal intercepts)	129	4	.000
Strict invariance (plus equal error variances)	157	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.999
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
socprob_fs	0.0	0.9	-0.5	4.3	22265

Share of cases with imputed missing values: 0.2%
(Equivalence of scores from robust MLMV: CD = .991)

List of scales (wave 0)

Scale: School reluctance

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	14239	3	.000

2)	**Root mean squared error (RMSEA)**		.000
	90% Confidence interval:	lower bound	.000
	90% Confidence interval:	upper bound	.000
	Probability RMSEA <= 0.05		1.000
3)	**Akaike's Information Criterion (AIC)**		245338
	Bayesian Information Criterion (BIC)		245410
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		1.000
	Tucker–Lewis Index (TLI)		1.000
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.000
	Coefficient of determination (CD)		.835

Reliability and Dimensionality

Ordinal Cronbach's Alpha .702

(Cronbach's alpha = .661)

McDonald's Omega .727

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.23
Factor 2	-.05
Factor 3	-.22

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
schoolav1	0.89	.007	0.88	0.91
schoolav2	0.67	.007	0.66	0.69
schoolav3	0.46	.006	0.45	0.47

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
schoolav1	3.1	1.8	1	6	22245
schoolav2	3.7	1.9	1	6	22248
schoolav3	2.2	1.5	1	6	22235

List of scales (wave 0)

Scale: School reluctance (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1451	9	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	99	2	.000
Strong invariance (plus equal intercepts)	981	2	.000
Strict invariance (plus equal error variances)	49	2	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	
Italian vs. German language version	

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	.994
Language: French/ Italian	.981

* **Note:** Due to sparse tables for the italian version of the scale, equivalence tests failed to converge and were reestimated with collapsed italian and french versions.

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
schoolav_fs	0.0	1.4	-2.0	2.6	22266

Share of cases with imputed missing values: 0.2%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Intrinsic achievement motivation — Full AES sample

Model and Fit Statistics

1) **Likelihood-ratio tests**

	chi2	df	p > chi2
Model vs. saturated	0	0	
Baseline vs. saturated	12995	3	.000

2) **Root mean squared error (RMSEA)** .000

90% Confidence interval:	lower bound	.000
90% Confidence interval:	upper bound	.000
Probability RMSEA <= 0.05		1.000

3) **Akaike's Information Criterion (AIC)** 152039

Bayesian Information Criterion (BIC) 152111

4) **Baseline comparison**

Comparative Fit Index (CFI)	1.000
Tucker–Lewis Index (TLI)	1.000

5) **Size of residuals**

Stand. root mean squared residual (SRMR)	.000
Coefficient of determination (CD)	.795

Reliability and Dimensionality

Ordinal Cronbach's Alpha .703

(Cronbach's alpha = .652)

McDonald's Omega .718

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.19
Factor 2	-.08
Factor 3	-.22

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
achmot2	0.54	.006	0.52	0.55
achmot4	0.62	.006	0.60	0.63
achmot6	0.86	.007	0.85	0.87

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
achmot2	3.0	0.8	1	4	22249
achmot4	2.8	0.8	1	4	22242
achmot6	2.6	0.9	1	4	22239

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
achmot2	1.16	-3.58	-1.45	1.12
achmot4	1.47	-3.30	-0.89	2.11
achmot6	2.88	-4.12	-0.77	3.70

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1286	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	14	4	.007
Strong invariance (plus equal intercepts)	956	4	.000
Strict invariance (plus equal error variances)	141	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.993
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.999
Language: French	.999
Language: Italian	.990

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
achmoti_fs	0.0	0.9	-2.2	1.8	22262

Share of cases with imputed missing values: 0.2%

(Equivalence of scores from robust MLMV: CD = .994)

(Equivalence of scores from two-step approach: CD = .982)

List of scales (wave 0)

Scale: Extrinsic achievement motivation

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	12774	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			148710
	Bayesian Information Criterion (BIC)			148782
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.792

Reliability and Dimensionality

Ordinal Cronbach's Alpha .648

(Cronbach's alpha = .589)

McDonald's Omega .690

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.14
Factor 2	-.04
Factor 3	-.22

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
achmot1	0.33	.007	0.32	0.34
achmot3	0.73	.009	0.72	0.75
achmot5	0.85	.009	0.83	0.86

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
achmot1	3.2	0.7	1	4	22263
achmot3	1.8	0.8	1	4	22239
achmot5	1.9	0.9	1	4	22235

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
achmot1	0.58	-3.66	-2.13	0.51
achmot3	2.18	-0.50	2.38	5.22
achmot5	2.49	-0.62	2.16	5.11

List of scales (wave 0)

Scale: Extrinsic achievement motivation (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1767	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	36	4	.000
Strong invariance (plus equal intercepts)	954	4	.000
Strict invariance (plus equal error variances)	211	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.982
French vs. Italian language version	.995
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	.979
Language: French	.961
Language: Italian	.993

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
achmote_fs	0.0	0.8	-1.3	2.3	22266

Share of cases with imputed missing values: 0.2%

(Equivalence of scores from robust MLMV: CD = .990)

(Equivalence of scores from two-step approach: CD = .981)

List of scales (wave 0)

Scale: Instrumental learning motivation (PISA2000) — Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	28969	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			144091
	Bayesian Information Criterion (BIC)			144163
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.865

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.848
(Cronbach's alpha = .796)	
McDonald's Omega	.850

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.81
Factor 2	-.10
Factor 3	-.14

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
insmot1	0.75	0.00	0.74	0.76
insmot2	0.79	0.00	0.78	0.80
insmot3	0.88	0.00	0.88	0.89

*** Note:** Replication of 'Insmot'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
insmot1	2.8	0.9	1	4	22246
insmot2	2.9	0.9	1	4	22220
insmot3	3.1	0.9	1	4	22220

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
insmot1	2.05	-3.82	-0.83	2.13
insmot2	2.35	-3.90	-1.28	1.70
insmot3	3.48	-6.32	-3.28	0.89

List of scales (wave 0)

Scale: Instrumental learning motivation (PISA2000) (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	347	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	29	4	.000
Strong invariance (plus equal intercepts)	136	4	.000
Strict invariance (plus equal error variances)	55	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.997
Italian vs. German language version	.994

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.982

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
insmot_fs	0.0	0.9	-2.2	1.4	22265

Share of cases with imputed missing values: 0.4%

(Equivalence of scores from robust MLMV: CD = .996)

(Equivalence of scores from two-step approach: CD = .978)

List of scales (wave 0)

Scale: Interest in reading

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	44643	3	.000

2)	**Root mean squared error (RMSEA)**		.000
	90% Confidence interval:	lower bound	.000
	90% Confidence interval:	upper bound	.000
	Probability RMSEA <= 0.05		1.000
3)	**Akaike's Information Criterion (AIC)**		153979
	Bayesian Information Criterion (BIC)		154051
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		1.000
	Tucker–Lewis Index (TLI)		1.000
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.000
	Coefficient of determination (CD)		.924

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.906
(Cronbach's alpha = .864)	
McDonald's Omega	.907

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.19
Factor 2	-.07
Factor 3	-.11

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
intrea1	0.86	.002	0.85	0.86
intrea2	0.94	.002	0.93	0.94
intrea3	0.83	.003	0.82	0.83

*** Note:** Replication of 'Intrea'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
intrea1	2.2	1.0	1	4	22180
intrea2	2.1	1.1	1	4	22178
intrea3	2.3	1.1	1	4	22165

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
intrea1	3.03	-1.81	0.96	3.55
intrea2	5.35	-1.65	2.08	5.65
intrea3	2.63	-1.67	0.17	2.61

List of scales (wave 0)

Scale: Interest in reading (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	732	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	94	4	.000
Strong invariance (plus equal intercepts)	560	4	.000
Strict invariance (plus equal error variances)	7	4	.155

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	1.000
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	.999
Language: French	.998
Language: Italian	.998

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
intrea_fs	0.0	0.9	-1.3	1.7	22200

Share of cases with imputed missing values: 0.3%
(Equivalence of scores from robust MLMV: CD = .997)
(Equivalence of scores from two-step approach: CD = .973)

List of scales (wave 0)

Scale: ICT interest

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	15929	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			69317
	Bayesian Information Criterion (BIC)			69383
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.884

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.855
(Cronbach's alpha = .797)	
McDonald's Omega	.860

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.88
Factor 2	-.09
Factor 3	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
ictmot2	0.69	.006	0.68	0.71
ictmot3	0.88	.004	0.87	0.89
ictmot4	0.87	.004	0.86	0.88

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
ictmot2	3.2	0.7	1	4	11068
ictmot3	2.4	1.0	1	4	11065
ictmot4	2.8	0.9	1	4	11060

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
ictmot2	1.77	-4.71	-2.62	0.94
ictmot3	3.41	-3.34	0.41	3.52
ictmot4	3.42	-4.79	-1.57	2.83

List of scales (wave 0)

Scale: ICT interest (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	408	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	69	4	.000
Strong invariance (plus equal intercepts)	95	4	.000
Strict invariance (plus equal error variances)	34	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.995
French vs. Italian language version	.997
Italian vs. German language version	.995

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.994
Language: Italian	.892

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
ictintr_fs	0.0	0.9	-2.1	1.6	11071

Share of cases with imputed missing values: 0.2%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of scores from two-step approach: CD = .992)

List of scales (wave 0)

Scale: Dispositional interest

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	1805	9	.000
	Baseline vs. saturated	31076	15	.000

2)	**Root mean squared error (RMSEA)**	.135
	90% Confidence interval: lower bound	.130
	90% Confidence interval: upper bound	.140
	Probability RMSEA <= 0.05	.000
3)	**Akaike's Information Criterion (AIC)**	137195
	Bayesian Information Criterion (BIC)	137326
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.942
	Tucker–Lewis Index (TLI)	.904
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.041
	Coefficient of determination (CD)	.888

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.875
(Cronbach's alpha = .836)	
McDonald's Omega	.876

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.19
Factor 2	.14
Factor 3	-.01
Factor 4	-.05
Factor 5	-.13
Factor 6	-.14

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
intsubj1	0.84	.004	0.83	0.85
intsubj2	0.65	.006	0.64	0.66
intsubj3	0.75	.005	0.74	0.76
intsubj4	0.66	.006	0.65	0.67
intsubj5	0.69	.006	0.68	0.71
intsubj6	0.80	.004	0.80	0.81

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
intsubj1	2.5	0.9	1	4	10889
intsubj2	3.2	0.7	1	4	10922
intsubj3	2.9	0.8	1	4	10845
intsubj4	2.6	0.9	1	4	10842
intsubj5	2.8	0.8	1	4	10905
intsubj6	2.4	1.0	1	4	10853

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
intsubj1	2.92	-3.37	-0.29	3.76
intsubj2	1.58	-4.54	-2.81	0.59
intsubj3	2.12	-4.06	-1.70	1.90
intsubj4	1.63	-2.34	-0.39	2.29
intsubj5	1.80	-3.88	-0.89	2.43
intsubj6	2.53	-2.10	0.31	3.26

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	885	54	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	91	10	.000
Strong invariance (plus equal intercepts)	332	10	.000
Strict invariance (plus equal error variances)	77	10	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.996
French vs. Italian language version	.995
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.999

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
intsubj_fs	0.0	0.9	-2.6	2.1	10949

Share of cases with imputed missing values: 1.6%
(Equivalence of scores from robust MLMV: CD = .999)
(Equivalence of scores from two-step approach: CD = .988)

List of scales (wave 0)

Scale: Identified motivation (mathematics)

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	45	2	.000
	Baseline vs. saturated	43936	6	.000
2)	**Root mean squared error (RMSEA)**			.044
	90% Confidence interval: lower bound			.034
	90% Confidence interval: upper bound			.056
	Probability RMSEA <= 0.05			.777
3)	**Akaike's Information Criterion (AIC)**			72033
	Bayesian Information Criterion (BIC)			72121
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.999
	Tucker–Lewis Index (TLI)			.997
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.004
	Coefficient of determination (CD)			.955

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.946
(Cronbach's alpha = .918)	
McDonald's Omega	.947

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.20
Factor 2	-.04
Factor 3	-.05
Factor 4	-.04

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
instrumot1	0.95	.001	0.94	0.95
instrumot2	0.93	.002	0.93	0.94
instrumot3	0.89	.002	0.88	0.89
instrumot4	0.85	.003	0.84	0.85

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
instrumot1	2.9	0.9	1	4	11018
instrumot2	2.9	0.9	1	4	11020
instrumot3	2.8	0.9	1	4	11030
instrumot4	2.9	0.9	1	4	11013

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
instrumot1	4.16	-7.00	-2.77	2.59
instrumot2	3.66	-5.86	-2.07	1.94
instrumot3	2.86	-5.38	-1.92	2.16
instrumot4	2.49	-5.04	-2.19	1.86

List of scales (wave 0)

Scale: Identified motivation (mathematics) (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	387	14	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	111	3	.000
Strong invariance (plus equal intercepts)	75	3	.000
Strict invariance (plus equal error variances)	135	3	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	
Italian vs. German language version	

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French/ Italian	1.000

* **Note:** Due to sparse tables for the italian version of the scale, equivalence tests failed to converge and were reestimated with collapsed italian and french versions.

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
instrumot_fs	-0.1	1.0	-2.4	1.5	11033

Share of cases with imputed missing values: 0.3%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of scores from two-step approach: CD = .985)

List of scales (wave 0)

Scale: External motivation regulation

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	687	2	.000
	Baseline vs. saturated	16452	6	.000
2)	**Root mean squared error (RMSEA)**			.177
	90% Confidence interval: lower bound			.166
	90% Confidence interval: upper bound			.188
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			100910
	Bayesian Information Criterion (BIC)			100998
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.958
	Tucker–Lewis Index (TLI)			.875
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.038
	Coefficient of determination (CD)			.844

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.820
(Cronbach's alpha = .764)	
McDonald's Omega	.826

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.06
Factor 2	.06
Factor 3	-.15
Factor 4	-.15

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
extreg2	0.76	.005	0.75	0.77
extreg3	0.81	.005	0.80	0.82
extreg4	0.58	.008	0.56	0.59
extreg5	0.78	.005	0.77	0.79

* **Note:** Items Extreg1 and Extreg6 Excluded to Improve Scale Quality

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
extreg2	1.9	0.9	1	4	10901
extreg3	2.0	0.9	1	4	10830
extreg4	2.4	0.9	1	4	10841
extreg5	1.8	0.9	1	4	10827

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
extreg2	2.11	-0.76	1.62	4.25
extreg3	2.55	-1.03	1.52	4.56
extreg4	1.28	-1.75	0.01	2.39
extreg5	2.34	-0.17	2.28	4.99

List of scales (wave 0)

Scale: External motivation regulation (continued)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	222	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	46	6	.000
Strong invariance (plus equal intercepts)	113	6	.000
Strict invariance (plus equal error variances)	35	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.990
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.998
Language: Italian	.997

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
extreg_fs	0.0	0.9	-1.4	2.5	10930

Share of cases with imputed missing values: 1.5%
(Equivalence of scores from robust MLMV: CD = .999)
(Equivalence of scores from two-step approach: CD = .977)

List of scales (wave 0)

Scale: Classroom participation

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	584	5	.000
	Baseline vs. saturated	28718	10	.000

2)	**Root mean squared error (RMSEA)**	.103
	90% Confidence interval: lower bound	.096
	90% Confidence interval: upper bound	.110
	Probability RMSEA <= 0.05	.000
3)	**Akaike's Information Criterion (AIC)**	97128
	Bayesian Information Criterion (BIC)	97238
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.980
	Tucker–Lewis Index (TLI)	.960
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.024
	Coefficient of determination (CD)	.890

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.888
(Cronbach's alpha = .848)	
McDonald's Omega	.888

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.95
Factor 2	.02
Factor 3	-.05
Factor 4	-.11
Factor 5	-.11

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
engage1	0.76	.005	0.75	0.77
engage2	0.83	.004	0.82	0.84
engage3	0.75	.005	0.74	0.76
engage4	0.80	.004	0.79	0.81
engage5	0.77	.005	0.76	0.78

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
engage1	2.9	0.8	1	4	10897
engage2	2.9	0.7	1	4	10852
engage3	3.0	0.7	1	4	10907
engage4	3.0	0.8	1	4	10898
engage5	2.8	0.8	1	4	10829

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
engage1	2.22	-4.53	-1.82	2.06
engage2	2.82	-5.44	-2.01	3.03
engage3	2.14	-4.97	-2.11	1.89
engage4	2.51	-5.30	-2.40	2.21
engage5	2.28	-4.28	-1.30	3.10

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	938	40	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	51	8	.000
Strong invariance (plus equal intercepts)	31	8	.000
Strict invariance (plus equal error variances)	149	8	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.997
French vs. Italian language version	.997
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.999

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
engage_fs	0.0	0.9	-2.7	1.9	10936

Share of cases with imputed missing values: 1.5%

(Equivalence of scores from robust MLMV: CD = .996)

(Equivalence of scores from two-step approach: CD = .984)

List of scales (wave 0)

Scale: Performance-approach goals (SELLMO)

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	620	2	.000
	Baseline vs. saturated	17637	6	.000

2)	**Root mean squared error (RMSEA)**	.171
	90% Confidence interval: lower bound	.159
	90% Confidence interval: upper bound	.182
	Probability RMSEA <= 0.05	.000
3)	**Akaike's Information Criterion (AIC)**	117025
	Bayesian Information Criterion (BIC)	117112
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.965
	Tucker–Lewis Index (TLI)	.895
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.040
	Coefficient of determination (CD)	.865

Reliability and Dimensionality

Ordinal Cronbach's Alpha .834
(Cronbach's alpha = .804)
McDonald's Omega .837

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.16
Factor 2	.05
Factor 3	-.15
Factor 4	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
approxgoals1	0.74	.006	0.73	0.75
approxgoals2	0.84	.004	0.83	0.84
approxgoals3	0.57	.008	0.55	0.58
approxgoals4	0.84	.004	0.83	0.85

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
approxgoals1	2.8	1.2	1	5	10608
approxgoals2	2.5	1.2	1	5	10478
approxgoals3	3.3	1.1	1	5	10596
approxgoals4	2.7	1.2	1	5	10474

List of scales (wave 0)

Scale: Performance-approach goals (SELLMO) (continued) Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	370	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	51	6	.000
Strong invariance (plus equal intercepts)	89	6	.000
Strict invariance (plus equal error variances)	76	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.988
Italian vs. German language version	.985

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.991

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
approxgoals_fs	0.0	0.8	-1.4	1.9	10628

Share of cases with imputed missing values: 1.8%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Learning goal orientation (SELLMO)

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	396	2	.000
	Baseline vs. saturated	16559	6	.000

2)	**Root mean squared error (RMSEA)**		.136
	90% Confidence interval:	lower bound	.125
	90% Confidence interval:	upper bound	.147
	Probability RMSEA <= 0.05		.000
3)	**Akaike's Information Criterion (AIC)**		113590
	Bayesian Information Criterion (BIC)		113677
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.976
	Tucker–Lewis Index (TLI)		.929
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.028
	Coefficient of determination (CD)		.841

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.839
(Cronbach's alpha = .808)	
McDonald's Omega	.839

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.15
Factor 2	-.01
Factor 3	-.15
Factor 4	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
learntarget1	0.74	.006	0.72	0.75
learntarget2	0.76	.006	0.75	0.77
learntarget3	0.73	.006	0.72	0.74
learntarget4	0.78	.005	0.77	0.79

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
learntarget1	3.3	1.1	1	5	10637
learntarget2	3.4	1.1	1	5	10481
learntarget3	3.3	1.1	1	5	10606
learntarget4	3.1	1.1	1	5	10485

List of scales (wave 0)

Scale: Learning goal orientation (SELLMO) (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	887	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	12	6	.072
Strong invariance (plus equal intercepts)	421	6	.000
Strict invariance (plus equal error variances)	254	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.999
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.997

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
learntarget_fs	0.0	0.7	-2.0	1.5	10649

Share of cases with imputed missing values: 1.8%

(Equivalence of scores from robust MLMV: CD = .998)

List of scales (wave 0)

Scale: Work avoidance (SELLMO)

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	370	2	.000
	Baseline vs. saturated	9625	6	.000
2)	**Root mean squared error (RMSEA)**			.131
	90% Confidence interval: lower bound			.120
	90% Confidence interval: upper bound			.143
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			122140
	Bayesian Information Criterion (BIC)			122227
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.962
	Tucker–Lewis Index (TLI)			.885
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.033
	Coefficient of determination (CD)			.761

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.747
(Cronbach's alpha = .712)	
McDonald's Omega	.750

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.59
Factor 2	-.02
Factor 3	-.09
Factor 4	-.22

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
avoidwork1	0.53	.009	0.51	0.54
avoidwork2	0.70	.007	0.68	0.71
avoidwork3	0.67	.008	0.66	0.69
avoidwork4	0.71	.007	0.70	0.72

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
avoidwork1	2.9	1.1	1	5	10615
avoidwork2	3.1	1.1	1	5	10483
avoidwork3	3.2	1.2	1	5	10599
avoidwork4	3.1	1.1	1	5	10480

List of scales (wave 0)

Scale: Work avoidance (SELLMO) (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	611	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	11	6	.087
Strong invariance (plus equal intercepts)	282	6	.000
Strict invariance (plus equal error variances)	170	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.989
Italian vs. German language version	.994

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.998
Language: Italian	.991

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
avoidwork_fs	0.0	0.5	-1.2	1.2	10637

Share of cases with imputed missing values: 1.8%
(Equivalence of scores from robust MLMV: CD = .996)

List of scales (wave 0)

Scale: Avoidance performance goals (SELLMO) Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	550	2	.000
	Baseline vs. saturated	20651	6	.000
2)	**Root mean squared error (RMSEA)**			.160
	90% Confidence interval: lower bound			.149
	90% Confidence interval: upper bound			.172
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			117023
	Bayesian Information Criterion (BIC)			117111
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.973
	Tucker–Lewis Index (TLI)			.920
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.027
	Coefficient of determination (CD)			.877

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.866
(Cronbach's alpha = .830)	
McDonald's Omega	.867

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.37
Factor 2	.01
Factor 3	-.09
Factor 4	-.14

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
avoidblame1	0.73	.005	0.72	0.74
avoidblame2	0.75	.005	0.74	0.76
avoidblame3	0.86	.004	0.85	0.87
avoidblame4	0.81	.005	0.80	0.81

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
avoidblame1	2.6	1.2	1	5	10594
avoidblame2	2.6	1.3	1	5	10496
avoidblame3	2.5	1.2	1	5	10604
avoidblame4	2.3	1.1	1	5	10509

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	378	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	19	6	.004
Strong invariance (plus equal intercepts)	120	6	.000
Strict invariance (plus equal error variances)	161	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.997
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
avoidblame_fs	0.0	0.8	-1.2	2.1	10642

Share of cases with imputed missing values: 1.9%

(Equivalence of scores from robust MLMV: CD = .998)

List of scales (wave 0)

Scale: Global self-esteem

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	20015	20	.000
	Baseline vs. saturated	64288	28	.000
2)	**Root mean squared error (RMSEA)**			.250
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			329588
	Bayesian Information Criterion (BIC)			329772
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.689
	Tucker-Lewis Index (TLI)			.564
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.147
	Coefficient of determination (CD)			.887

Reliability and Dimensionality

Ordinal Cronbach's Alpha .859
(Cronbach's alpha = .820)
McDonald's Omega .852

Test of (one-)dimensionality (parallel analysis)
Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	3.56
factor 2	1.12
factor 3	.07
factor 4	-.05
factor 5	-.09
factor 6	-.10
factor 7	-.12
factor 8	-.13

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
sele1	0.63	0.01	0.62	0.64
sele2	0.51	0.01	0.49	0.52
sele3	0.44	0.01	0.43	0.46
sele4	0.49	0.01	0.48	0.51
seld1	0.85	0.00	0.84	0.85
seld3	0.75	0.00	0.74	0.75
seld4	0.65	0.01	0.64	0.66
seld5	0.80	0.00	0.79	0.81

*** Note:** Reversed categories for all seld-items

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
sele1	4.0	0.9	1	5	15991
sele2	4.1	0.8	1	5	15961
sele3	3.9	0.8	1	5	15957
sele4	3.8	1.0	1	5	15946
seld1	3.8	1.2	1	5	15972
seld3	3.2	1.2	1	5	15953
seld4	3.2	1.3	1	5	15902
seld5	4.0	1.2	1	5	15943

List of scales (wave 0)

Scale: Global self-esteem (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	5550	88	.000	693	44	.000	136	44	.000

Tests of measurement invariance across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	85	14	.000	27	7	.000	38	7	.000
Strong invariance (plus equal intercepts)	3216	14	.000	618	7	.000	42	7	.000
Strict invariance (plus equal error variances)	415	14	.000	205	7	.000	25	7	.001

Configural factor similarity across ...

Tucker's congruence coefficient

Survey languages	TCC	Survey settings	TCC	Survey modes	TCC
German vs. French	.999	classroom vs. unproctored	.999	web vs. PAP	.991
French vs. Italian	.998				
Italian vs. German	.996				

Factor score equivalence: group specific vs. invariant models for ...

Coefficient of determination

Survey languages	CD	Survey settings	CD	Survey modes	CD
German	1.000	classroom	1.000	web	1.000
French	.994	unproctored	.998	PAP	.985
Italian	.989				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
sel_fs	0.0	0.5	-1.8	0.8	16003

Share of cases with imputed missing values: 1.2%

(Equivalence of scores from Robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Global self-esteem (shortened)

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	17789	14	.000
	Baseline vs. saturated	55337	21	.000
2)	**Root mean squared error (RMSEA)**			.282
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			283054
	Bayesian Information Criterion (BIC)			283215
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.679
	Tucker-Lewis Index (TLI)			.518
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.133
	Coefficient of determination (CD)			.860

Reliability and Dimensionality

Ordinal Cronbach's Alpha .852
(Cronbach's alpha = .809)
McDonald's Omega .852

Test of (one-)dimensionality (parallel analysis)
Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	3.24
factor 2	.97
factor 3	-.01
factor 4	-.06
factor 5	-.11
factor 6	-.12
factor 7	-.14

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf.	interval]
sele1	0.77	0.00	0.76	0.78
sele2	0.72	0.01	0.71	0.73
sele3	0.67	0.01	0.66	0.68
sele4	0.68	0.01	0.67	0.69
seld1	0.66	0.01	0.64	0.67
seld3	0.56	0.01	0.55	0.58
seld5	0.63	0.01	0.62	0.64

* **Note:** Reversed categories for all seld-items

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
sele1	4.0	0.9	1	5	15991
sele2	4.1	0.8	1	5	15961
sele3	3.9	0.8	1	5	15957
sele4	3.8	1.0	1	5	15946
seld1	3.8	1.2	1	5	15972
seld3	3.2	1.2	1	5	15953
seld5	4.0	1.2	1	5	15943

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	4643	70	.000	628	35	.000	125	35	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	130	12	.000	40	6	.000	12	6	.069
Strong invariance (plus equal intercepts)	1838	12	.000	589	6	.000	52	6	.000
Strict invariance (plus equal error variances)	320	12	.000	142	6	.000	15	6	.017

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.996	classroom vs. unproctored	.996	web vs. PAP	.996
	French vs. Italian	.983				
	Italian vs. German	.966				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	.999	classroom	.999	web	1.000
	French	.997	unproctored	.991	PAP	.997
	Italian	.826				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
sel_m_fs	0.0	0.6	-2.7	1.0	16003

Share of cases with imputed missing values: 1.0%

(Equivalence of Scores from Robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Positive global self-esteem

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	329	2	.000
	Baseline vs. saturated	26567	6	.000

2)	**Root mean squared error (RMSEA)**		.101
	90% Confidence interval:	lower bound	.092
	90% Confidence interval:	upper bound	.110
	Probability RMSEA <= 0.05		.000
3)	**Akaike's Information Criterion (AIC)**		140371
	Bayesian Information Criterion (BIC)		140463
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.988
	Tucker-Lewis Index (TLI)		.963
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.018
	Coefficient of determination (CD)		.856

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.848
(Cronbach's alpha = .801)	
McDonald's Omega	.849

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.21
factor 2	-.06
factor 3	-.07
factor 4	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
sele1	0.72	0.00	0.71	0.73
sele2	0.83	0.00	0.82	0.83
sele3	0.78	0.00	0.78	0.79
sele4	0.72	0.00	0.71	0.73

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
sele1	4.0	0.9	1	5	15991
sele2	4.1	0.8	1	5	15961
sele3	3.9	0.8	1	5	15957
sele4	3.8	1.0	1	5	15946

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	1803	28	.000	346	14	.000	35	14	.002

Tests of measurement invariance across ...	**Survey languages**			**Survey settings**			**Survey modes**		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	21	6	.002	11	3	.013	1	3	.769
Strong invariance (plus equal intercepts)	1214	6	.000	140	3	.000	8	3	.052
Strict invariance (plus equal error variances)	216	6	.000	123	3	.000	10	3	.017

Configural factor similarity across ...	**Survey languages**		**Survey settings**		**Survey modes**	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	1.000	classroom vs. unproctored	1.000	web vs. PAP	1.000
	French vs. Italian	.998				
	Italian vs. German	.997				

Factor score equivalence: group specific vs. invariant models for ...	**Survey languages**		**Survey settings**		**Survey modes**	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	.998	unproctored	1.000	PAP	1.000
	Italian	.992				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
sele_fs	0.0	0.6	-2.5	0.9	15997

Share of cases with imputed missing values: 0.6%
(Equivalence of scores from robust MLMV: CD = .996)

List of scales (wave 0)

Scale: Negative global self-esteem

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	712	2	.000
	Baseline vs. saturated	31810	6	.000
2)	**Root mean squared error (RMSEA)**			.149
	90% Confidence interval: lower bound			.140
	90% Confidence interval: upper bound			.158
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			175983
	Bayesian Information Criterion (BIC)			176075
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.978
	Tucker-Lewis Index (TLI)			.933
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.028
	Coefficient of determination (CD)			.887

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.866
(Cronbach's alpha = .824)	
McDonald's Omega	.868

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.39
factor 2	.02
factor 3	-.13
factor 4	-.12

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
seld1	0.88	0.00	0.88	0.89
seld3	0.79	0.00	0.78	0.80
seld4	0.67	0.01	0.66	0.68
seld5	0.80	0.00	0.80	0.81

*** Note:** Reversed Item Categories

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
seld1	3.8	1.2	1	5	15972
seld3	3.2	1.2	1	5	15953
seld4	3.2	1.3	1	5	15902
seld5	4.0	1.2	1	5	15943

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	4554	28	.000	140	14	.000	59	14	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	107	6	.000	4	3	.235	7	3	.064
Strong invariance (plus equal intercepts)	2496	6	.000	86	3	.000	27	3	.000
Strict invariance (plus equal error variances)	355	6	.000	1	3	.707	7	3	.089

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.997	classroom vs. unproctored	1.000	web vs. PAP	.998
	French vs. Italian	1.000				
	Italian vs. German	.998				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	.990	unproctored	1.000	PAP	.999
	Italian	.980				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
seld_fs	0.0	1.0	-2.6	1.3	15995

Share of cases with imputed missing values: 0.9%

(Equivalence of scores from robust MLMV: CD = .993)

List of scales (wave 0)

Scale: Negative global self-esteem (shortened)

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	23184	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			130616
	Bayesian Information Criterion (BIC)			130685
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.885

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.863
(Cronbach's alpha = .816)	
McDonald's Omega	.865

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.90
factor 2	-.08
factor 3	-.14

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
seld1	0.91	0.00	0.90	0.91
seld3	0.80	0.00	0.79	0.80
seld5	0.77	0.00	0.76	0.78

*** Note:** Reversed Item Categories

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
seld1	3.8	1.2	1	5	15972
seld3	3.2	1.2	1	5	15953
seld5	4.0	1.2	1	5	15943

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	2872	18	.000	104	9	.000	53	9	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	61	4	.000	1	2	.749	6	2	.061
Strong invariance (plus equal intercepts)	1218	4	.000	62	2	.000	26	2	.000
Strict invariance (plus equal error variances)	27	4	.000	1	2	.511	5	2	.087

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.998	classroom vs. unproctored	1.000	web vs. PAP	.998
	French vs. Italian	1.000				
	Italian vs. German	.998				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	.999	classroom	1.000	web	1.000
	French	.989	unproctored	1.000	PAP	.997
	Italian	.980				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
seld_m_fs	0.0	1.0	-2.6	1.2	15994

Share of cases with imputed missing values: 0.5%
(Equivalence of Scores from Robust MLMV: CD = .999)

List of scales (wave 0)

Scale: General perceived self-efficacy scale (GSES)

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	63	2	.000
	Baseline vs. saturated	23581	6	.000
2)	**Root mean squared error (RMSEA)**			.044
	90% Confidence interval: lower bound			.035
	90% Confidence interval: upper bound			.053
	Probability RMSEA <= 0.05			.847
3)	**Akaike's Information Criterion (AIC)**			104477
	Bayesian Information Criterion (BIC)			104569
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.997
	Tucker-Lewis Index (TLI)			.992
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.009
	Coefficient of determination (CD)			.836

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.835
(Cronbach's alpha = .772)	
McDonald's Omega	.835

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.10
factor 2	-.08
factor 3	-.12
factor 4	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
seef1	0.73	0.00	0.72	0.74
seef2	0.77	0.00	0.76	0.78
seef3	0.76	0.00	0.75	0.77
seef4	0.73	0.00	0.72	0.74

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
seef1	3.1	0.6	1	4	15941
seef2	3.1	0.7	1	4	15928
seef3	2.8	0.7	1	4	15916
seef4	3.0	0.7	1	4	15923

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
seef1	2.04	-6.05	-3.17	2.22
seef2	2.28	-6.20	-2.91	1.82
seef3	2.14	-5.09	-1.43	2.66
seef4	2.03	-5.56	-2.00	2.27

List of scales (wave 0)

Scale: General perceived self-efficacy scale (GSES) (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	1049	28	.000	104	14	.000	24	14	.044

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	47	6	.000	1	3	.763	4	3	.252
Strong invariance (plus equal intercepts)	448	6	.000	10	3	.018	2	3	.652
Strict invariance (plus equal error variances)	230	6	.000	12	3	.008	4	3	.303

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.998	classroom vs. unproctored	1.000	web vs. PAP	.999
	French vs. Italian	.995				
	Italian vs. German	.996				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	1.000
	French	.997	unproctored	1.000	PAP	.999
	Italian	.993				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
seef_fs	0.0	0.9	-3.0	1.8	15951

Share of cases with imputed missing values: 0.4%

(Equivalence of scores from robust MLMV: CD = .996)

(Equivalence of Scores from Two-Step-Approach: CD = .989)

List of scales (wave 0)

Scale: Academic self-efficacy

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	32752	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			179405
	Bayesian Information Criterion (BIC)			179477
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.874

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.868
(Cronbach's alpha = .836)	
McDonald's Omega	.869

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.92
Factor 2	-.11
Factor 3	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
acaself1	0.81	.003	0.80	0.81
acaself2	0.87	.003	0.87	0.88
acaself3	0.81	.003	0.80	0.81

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
acaself1	4.7	1.1	1	6	22256
acaself2	4.1	1.2	1	6	22248
acaself3	4.3	1.2	1	6	22252

List of scales (wave 0)

Scale: Academic self-efficacy (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	774	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	77	4	.000
Strong invariance (plus equal intercepts)	250	4	.000
Strict invariance (plus equal error variances)	318	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.998
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.989

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
acaself_fs	0.0	0.8	-2.7	1.4	22264

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Academic self-concept (PISA2000) — Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	31794	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			111791
	Bayesian Information Criterion (BIC)			111863
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.884

Reliability and Dimensionality

Ordinal Cronbach's Alpha .856
(Cronbach's alpha = .795)
McDonald's Omega .860

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.89
Factor 2	-.08
Factor 3	-.14

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf.	interval]
scacad1	0.70	.004	0.70	0.71
scacad2	0.89	.003	0.89	0.90
scacad3	0.85	.003	0.84	0.86

*** Note:** Replication of 'Scacad'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
scacad1	2.9	0.7	1	4	22202
scacad2	2.9	0.7	1	4	22175
scacad3	2.9	0.7	1	4	22168

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
scacad1	1.87	-4.54	-1.94	2.37
scacad2	3.96	-7.57	-2.86	3.92
scacad3	3.05	-6.36	-2.61	3.41

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1571	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	76	4	.000
Strong invariance (plus equal intercepts)	768	4	.000
Strict invariance (plus equal error variances)	427	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.999
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.999
Language: French	.987
Language: Italian	.996

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
scacad_fs	0.0	0.9	-2.5	1.7	22210

Share of cases with imputed missing values: 0.3%
(Equivalence of scores from robust MLMV: CD = .997)
(Equivalence of scores from two-step approach: CD = .986)

List of scales (wave 0)

Scale: Verbal self-concept (PISA2000) — Full AES sample

Model and Fit Statistics

1) **Likelihood-ratio tests**

	chi2	df	p > chi2
Model vs. saturated	0	0	
Baseline vs. saturated	32226	3	.000

2) **Root mean squared error (RMSEA)** .000

90% Confidence interval:	lower bound	.000
90% Confidence interval:	upper bound	.000
Probability RMSEA <= 0.05		1.000

3) **Akaike's Information Criterion (AIC)** 128063

Bayesian Information Criterion (BIC) 128135

4) **Baseline comparison**

Comparative Fit Index (CFI)	1.000
Tucker–Lewis Index (TLI)	1.000

5) **Size of residuals**

Stand. root mean squared residual (SRMR)	.000
Coefficient of determination (CD)	.888

Reliability and Dimensionality

Ordinal Cronbach's Alpha .856

(Cronbach's alpha = .795)

McDonald's Omega .861

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.90
Factor 2	-.08
Factor 3	-.14

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
scverb1 **	0.70	0.00	0.69	0.70
scverb2	0.90	0.00	0.89	0.90
scverb3	0.86	0.00	0.85	0.86

*** Note:** Replication of 'Scverb'-Scale from TREE1 / PISA2000

**** Note:** Reversed Categories for Item Scverb1

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
scverb1 **	3.2	0.8	1	4	22196
scverb2	2.8	0.8	1	4	22173
scverb3	2.9	0.8	1	4	22171

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
scverb1	1.84	-4.49	-2.24	0.34
scverb2	3.52	-6.01	-1.79	3.39
scverb3	2.89	-5.94	-2.37	2.79

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	621	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	30	4	.000
Strong invariance (plus equal intercepts)	58	4	.000
Strict invariance (plus equal error variances)	215	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.989
Italian vs. German language version	.986

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.998
Language: Italian	.998

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
scverb_fs	0.0	0.9	-2.4	1.6	22205

Share of cases with imputed missing values: 0.3%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of scores from two-step approach: CD = .988)

List of scales (wave 0)

Scale: Maths self-concept [PISA 2000]

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	57824	3	.000

2)	**Root mean squared error (RMSEA)**	.000
	90% Confidence interval: lower bound	.000
	90% Confidence interval: upper bound	.000
	Probability RMSEA <= 0.05	1.000
3)	**Akaike's Information Criterion (AIC)**	134733
	Bayesian Information Criterion (BIC)	134805
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	1.000
	Tucker–Lewis Index (TLI)	1.000
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.000
	Coefficient of determination (CD)	.980

Reliability and Dimensionality

Ordinal Cronbach's Alpha .927
(Cronbach's alpha = .888)
McDonald's Omega .930

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.38
Factor 2	-.01
Factor 3	-.08

Standardized factor loadings

Indicators *	Coef.	(SE)	[95% Conf. interval]	
matcon1	0.90	.002	0.90	0.90
matcon2	0.99	.001	0.99	0.99
matcon3	0.82	.002	0.81	0.82

* **Note:** Replication of 'Matcon'-Scale from TREE1 / PISA2000

Item descriptives

Indicators *	Mean	Std. dev.	Min.	Max.	Valid Obs.
matcon1	2.7	0.9	1	4	22183
matcon2	2.4	1.1	1	4	22187
matcon3	2.4	1.0	1	4	22180

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
matcon1	3.38	-4.50	-1.06	2.95
matcon2	4.96	-3.20	0.21	4.25
matcon3	2.40	-2.30	0.21	2.53

List of scales (wave 0)

Scale: Maths self-concept [PISA 2000] (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	937	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	335	4	.000
Strong invariance (plus equal intercepts)	47	4	.000
Strict invariance (plus equal error variances)	241	2	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.997
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	1.000

*** Note:** Language-specific models do not converge and the related invariance tests and indices may not be calculated unless the error variance of item matcon2 is constrained to zero.

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
matcon_fs	0.0	1.0	-1.7	1.6	22193

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .967)

(Equivalence of scores from two-step approach: CD = .899)

List of scales (wave 0)

Scale: ICT self-concept

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	20861	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			68148
	Bayesian Information Criterion (BIC)			68214
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.912

Reliability and Dimensionality

Ordinal Cronbach's Alpha .896
(Cronbach's alpha = .849)
McDonald's Omega .898

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.12
Factor 2	-.08
Factor 3	-.10

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
ictmot6	0.78	.004	0.77	0.79
ictmot7	0.90	.003	0.89	0.90
ictmot8	0.91	.003	0.90	0.91

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
ictmot6	2.9	0.9	1	4	11064
ictmot7	2.2	0.9	1	4	11057
ictmot8	2.4	0.9	1	4	11058

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
ictmot6	2.31	-4.15	-1.43	1.80
ictmot7	3.82	-2.56	1.99	5.06
ictmot8	4.06	-3.74	0.04	4.72

List of scales (wave 0)

Scale: ICT self-concept (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	628	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	82	4	.000
Strong invariance (plus equal intercepts)	47	4	.000
Strict invariance (plus equal error variances)	170	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.996
French vs. Italian language version	.987
Italian vs. German language version	.997

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.996

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
ictabil_fs	0.0	0.9	-1.8	1.8	11067

Share of cases with imputed missing values: 0.2%

(Equivalence of scores from robust MLMV: CD = .997)

(Equivalence of scores from two-step approach: CD = .989)

List of scales (wave 0)

Scale: Specific self-efficacy: numeracy

Full AES sample

Model and Fit Statistics

1) **Likelihood-ratio tests**

	chi2	df	p > chi2
Model vs. saturated	536	2	.000
Baseline vs. saturated	36814	6	.000

2) **Root mean squared error (RMSEA)** .110

90% Confidence interval:	lower bound	.103
90% Confidence interval:	upper bound	.118
Probability RMSEA <= 0.05		.000

3) **Akaike's Information Criterion (AIC)** 196455

Bayesian Information Criterion (BIC) 196551

4) **Baseline comparison**

Comparative Fit Index (CFI)	.985
Tucker–Lewis Index (TLI)	.956

5) **Size of residuals**

Stand. root mean squared residual (SRMR)	.020
Coefficient of determination (CD)	.854

Reliability and Dimensionality

Ordinal Cronbach's Alpha .851

(Cronbach's alpha = .831)

McDonald's Omega .852

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.23
Factor 2	-.05
Factor 3	-.08
Factor 4	-.16

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
selfeff01	0.77	.004	0.76	0.77
selfeff02	0.77	.004	0.76	0.78
selfeff03	0.80	.003	0.79	0.81
selfeff04	0.73	.004	0.72	0.74

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
selfeff01	3.3	0.9	1	4	21801
selfeff02	3.0	0.9	1	4	21827
selfeff03	2.8	0.9	1	4	10734
selfeff04	2.7	0.9	1	4	10755

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
selfeff01	2.35	-4.76	-2.62	-0.16
selfeff02	2.38	-4.13	-1.77	1.07
selfeff03	3.03	-5.40	-1.83	2.94
selfeff04	2.27	-4.13	-1.09	2.49

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	651	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	72	6	.000
Strong invariance (plus equal intercepts)	85	6	.000
Strict invariance (plus equal error variances)	33	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.998
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.998
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
selfeffa_fs	0.0	0.9	-2.4	1.6	21881

Share of cases with imputed missing values: 51.2%

(Equivalence of scores from robust MLMV: CD = .995)

(Equivalence of scores from two-step approach: CD = .976)

List of scales (wave 0)

Scale: Specific self-efficacy: algebra

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	3889	2	.000
	Baseline vs. saturated	92426	6	.000

2)	**Root mean squared error (RMSEA)**		.298
	90% Confidence interval:	lower bound	.290
	90% Confidence interval:	upper bound	.306
	Probability RMSEA <= 0.05		.000
3)	**Akaike's Information Criterion (AIC)**		147967
	Bayesian Information Criterion (BIC)		148063
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.958
	Tucker–Lewis Index (TLI)		.874
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.026
	Coefficient of determination (CD)		.957

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.947
(Cronbach's alpha = .926)	
McDonald's Omega	.948

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.24
Factor 2	.07
Factor 3	-.06
Factor 4	-.06

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
selfeff05	0.86	.002	0.86	0.87
selfeff06	0.95	.001	0.95	0.96
selfeff07	0.88	.002	0.88	0.89
selfeff08	0.92	.001	0.92	0.93

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
selfeff05	3.3	0.9	1	4	21809
selfeff06	3.0	1.0	1	4	21794
selfeff07	2.8	1.0	1	4	10747
selfeff08	3.2	0.9	1	4	10730

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
selfeff05	3.39	-5.99	-3.58	-0.95
selfeff06	8.35	-11.55	-5.35	1.58
selfeff07	4.65	-6.43	-2.51	1.99
selfeff08	5.99	-9.89	-5.56	-0.57

List of scales (wave 0)

Scale: Specific self-efficacy: algebra (continued) — Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	506	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	17	6	.010
Strong invariance (plus equal intercepts)	116	6	.000
Strict invariance (plus equal error variances)	238	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	1.000
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.998

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
selfeffb_fs	-0.1	0.9	-2.2	1.1	21872

Share of cases with imputed missing values: 51.2%

(Equivalence of scores from robust MLMV: CD = .998)

(Equivalence of scores from two-step approach: CD = .957)

List of scales (wave 0)

Scale: Specific self-efficacy: geometry

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	229	2	.000
	Baseline vs. saturated	30977	6	.000
2)	**Root mean squared error (RMSEA)**			.072
	90% Confidence interval: lower bound			.064
	90% Confidence interval: upper bound			.080
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			203347
	Bayesian Information Criterion (BIC)			203443
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.993
	Tucker–Lewis Index (TLI)			.978
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.015
	Coefficient of determination (CD)			.836

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.823
(Cronbach's alpha = .803)	
McDonald's Omega	.825

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.05
Factor 2	-.07
Factor 3	-.09
Factor 4	-.16

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
selfeff09	0.81	.004	0.80	0.81
selfeff10	0.76	.004	0.75	0.76
selfeff11	0.75	.004	0.74	0.75
selfeff12	0.63	.005	0.62	0.64

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
selfeff09	3.3	0.9	1	4	10752
selfeff10	3.2	0.9	1	4	21783
selfeff11	3.0	1.0	1	4	21802
selfeff12	2.6	0.9	1	4	10751

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
selfeff09	3.22	-6.78	-3.69	-0.03
selfeff10	2.24	-4.55	-2.29	0.17
selfeff11	2.15	-3.88	-1.49	0.85
selfeff12	1.75	-3.32	-0.62	2.77

List of scales (wave 0)

Scale: Specific self-efficacy: geometry (continued) Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	3499	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	59	6	.000
Strong invariance (plus equal intercepts)	2400	6	.000
Strict invariance (plus equal error variances)	320	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.997
Italian vs. German language version	.993

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.993
Language: Italian	.988

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
selfeffc_fs	0.0	0.9	-2.5	1.5	21875

Share of cases with imputed missing values: 51.3%
(Equivalence of scores from robust MLMV: CD = .995)
(Equivalence of scores from two-step approach: CD = .965)

List of scales (wave 0)

Scale: Specific self-efficacy: probability

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	1326	2	.000
	Baseline vs. saturated	63299	6	.000
2)	**Root mean squared error (RMSEA)**			.174
	90% Confidence interval: lower bound			.166
	90% Confidence interval: upper bound			.182
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			178726
	Bayesian Information Criterion (BIC)			178821
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.979
	Tucker–Lewis Index (TLI)			.937
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.022
	Coefficient of determination (CD)			.919

Reliability and Dimensionality

Ordinal Cronbach's Alpha .917

(Cronbach's alpha = .907)

McDonald's Omega .917

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.86
Factor 2	.01
Factor 3	-.09
Factor 4	-.10

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
selfeff13	0.87	.002	0.86	0.87
selfeff14	0.84	.002	0.83	0.84
selfeff15	0.89	.002	0.88	0.89
selfeff16	0.83	.003	0.83	0.84

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
selfeff13	2.7	1.0	1	4	21778
selfeff14	2.6	1.0	1	4	10754
selfeff15	2.8	0.9	1	4	21776
selfeff16	2.5	0.9	1	4	10751

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
selfeff13	3.46	-4.44	-0.85	2.41
selfeff14	3.65	-4.88	-0.67	3.58
selfeff15	3.96	-5.27	-1.24	2.74
selfeff16	3.51	-4.69	-0.45	3.96

List of scales (wave 0)

Scale: Specific self-efficacy: probability (continued)

Full AES sample

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	118	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	11	6	.102
Strong invariance (plus equal intercepts)	42	6	.000
Strict invariance (plus equal error variances)	21	6	.002

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	1.000
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
selfeffd_fs	0.0	0.9	-2.0	1.7	21858

Share of cases with imputed missing values: 51.2%

(Equivalence of scores from robust MLMV: CD = .997)

(Equivalence of scores from two-step approach: CD = .986)

List of scales (wave 0)

Scale: Mathematics anxiety

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	1904	5	.000
	Baseline vs. saturated	37885	10	.000

2)	**Root mean squared error (RMSEA)**	.186
	90% Confidence interval: lower bound	.179
	90% Confidence interval: upper bound	.193
	Probability RMSEA <= 0.05	.000
3)	**Akaike's Information Criterion (AIC)**	114426
	Bayesian Information Criterion (BIC)	114535
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.950
	Tucker–Lewis Index (TLI)	.900
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.035
	Coefficient of determination (CD)	.916

Reliability and Dimensionality

Ordinal Cronbach's Alpha .914
(Cronbach's alpha = .877)
McDonald's Omega .914

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.35
Factor 2	.10
Factor 3	-.03
Factor 4	-.10
Factor 5	-.12

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
anxmath1	0.83	.004	0.82	0.84
anxmath2	0.79	.004	0.79	0.80
anxmath3	0.84	.004	0.83	0.85
anxmath4	0.80	.004	0.79	0.81
anxmath5	0.86	.003	0.85	0.86

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
anxmath1	2.4	1.0	1	4	10999
anxmath2	1.9	0.9	1	4	10996
anxmath3	1.8	0.9	1	4	10992
anxmath4	2.5	1.0	1	4	10995
anxmath5	2.1	1.0	1	4	10994

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
anxmath1	2.83	-2.61	0.30	3.40
anxmath2	2.48	-0.85	1.87	4.52
anxmath3	2.94	-0.26	2.70	5.39
anxmath4	2.50	-2.49	-0.24	2.32
anxmath5	3.11	-1.59	1.60	4.41

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1137	40	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	141	8	.000
Strong invariance (plus equal intercepts)	502	8	.000
Strict invariance (plus equal error variances)	151	8	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.995
Italian vs. German language version	.988

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.980

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
anxmath_fs	0.0	0.9	-1.6	2.3	11005

Share of cases with imputed missing values: 0.2%
(Equivalence of scores from robust MLMV: CD = .999)
(Equivalence of scores from two-step approach: CD = .976)

List of scales (wave 0)

Scale: Mathematics boredom

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	689	2	.000
	Baseline vs. saturated	20215	6	.000
2)	**Root mean squared error (RMSEA)**			.178
	90% Confidence interval: lower bound			.167
	90% Confidence interval: upper bound			.189
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			125128
	Bayesian Information Criterion (BIC)			125216
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.966
	Tucker–Lewis Index (TLI)			.898
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.032
	Coefficient of determination (CD)			.863

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.863
(Cronbach's alpha = .831)	
McDonald's Omega	.863

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.34
Factor 2	.02
Factor 3	-.11
Factor 4	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
boredom1	0.78	.005	0.77	0.79
boredom2	0.78	.005	0.77	0.79
boredom3	0.80	.005	0.79	0.81
boredom4	0.77	.005	0.76	0.78

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
boredom1	2.9	1.3	1	5	10877
boredom2	2.6	1.2	1	5	10834
boredom3	2.5	1.3	1	5	10813
boredom4	3.0	1.3	1	5	10877

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	815	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	15	6	.022
Strong invariance (plus equal intercepts)	599	6	.000
Strict invariance (plus equal error variances)	166	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.997
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.995

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
boredom_fs	0.0	0.9	-1.5	1.9	10902

Share of cases with imputed missing values: 1.1%
(Equivalence of scores from robust MLMV: CD = .998)

List of scales (wave 0)

Scale: Mathematics anger

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	79	2	.000
	Baseline vs. saturated	27251	6	.000
2)	**Root mean squared error (RMSEA)**			.059
	90% Confidence interval: lower bound			.049
	90% Confidence interval: upper bound			.071
	Probability RMSEA <= 0.05			.073
3)	**Akaike's Information Criterion (AIC)**			120644
	Bayesian Information Criterion (BIC)			120732
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.997
	Tucker–Lewis Index (TLI)			.992
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.010
	Coefficient of determination (CD)			.915

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.895
(Cronbach's alpha = .865)	
McDonald's Omega	.897

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.66
Factor 2	-.05
Factor 3	-.08
Factor 4	-.09

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
anger1	0.70	.005	0.69	0.71
anger2	0.89	.003	0.89	0.90
anger3	0.89	.003	0.88	0.89
anger4	0.82	.004	0.82	0.83

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
anger1	2.6	1.2	1	5	10891
anger2	2.4	1.3	1	5	10815
anger3	2.5	1.3	1	5	10810
anger4	2.5	1.4	1	5	10869

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1045	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	52	6	.000
Strong invariance (plus equal intercepts)	264	6	.000
Strict invariance (plus equal error variances)	48	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.997
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.996

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
anger_fs	0.0	0.9	-1.4	2.1	10902

Share of cases with imputed missing values: 1.1%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Mathematics enjoyment

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	191	2	.000
	Baseline vs. saturated	23069	6	.000
2)	**Root mean squared error (RMSEA)**			.093
	90% Confidence interval: lower bound			.082
	90% Confidence interval: upper bound			.104
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			114281
	Bayesian Information Criterion (BIC)			114369
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.992
	Tucker–Lewis Index (TLI)			.975
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.014
	Coefficient of determination (CD)			.892

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.877
(Cronbach's alpha = .845)	
McDonald's Omega	.879

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.47
Factor 2	-.04
Factor 3	-.09
Factor 4	-.11

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
enjoymath1	0.86	.004	0.86	0.87
enjoymath2	0.86	.004	0.86	0.87
enjoymath3	0.73	.005	0.72	0.74
enjoymath4	0.75	.005	0.74	0.76

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
enjoymath1	2.5	1.2	1	5	10880
enjoymath2	2.5	1.2	1	5	10830
enjoymath3	2.3	1.2	1	5	10882
enjoymath4	2.3	1.1	1	5	10823

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	333	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	44	6	.000
Strong invariance (plus equal intercepts)	152	6	.000
Strict invariance (plus equal error variances)	40	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	1.000
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.998

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
enjoymath_fs	0.0	0.9	-1.4	2.5	10907

Share of cases with imputed missing values: 1.0%
(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Perseverance

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	18182	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			168695
	Bayesian Information Criterion (BIC)			168767
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.825

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.767
(Cronbach's alpha = .731)	
McDonald's Omega	.775

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.43
Factor 2	-.09
Factor 3	-.20

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
persev1	0.67	.005	0.66	0.68
persev2	0.87	.005	0.86	0.88
persev3	0.64	.005	0.63	0.65

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
persev1	3.5	0.9	1	5	22268
persev2	3.4	1.0	1	5	22269
persev3	2.9	1.0	1	5	22265

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	2678	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	79	4	.000
Strong invariance (plus equal intercepts)	1498	4	.000
Strict invariance (plus equal error variances)	207	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.997
French vs. Italian language version	.999
Italian vs. German language version	.994

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	.998
Language: French	.990
Language: Italian	.989

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
persev_fs	0.0	0.5	-1.5	1.1	22280

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Effort: learning (PISA2000)

Full AES sample

Composite descriptives	Variable name	Mean	Std. dev.	Min.	Max.	Obs.
	effper_comp	2.8	0.8	1	4	22265
Share of cases with imputed missing values:		0.2%				

Item descriptives	Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
	effper1 *	2.7	0.8	1	4	22243
	effper4 *	2.9	0.9	1	4	22249

* **Note:** Original items from TREE1 / PISA2000

List of scales (wave 0)

Big five inventory

Baseline survey sample

Composite descriptives	Variable name	Mean	Std. dev.	Min.	Max.	Obs.
Big five: extraversion						
	big5_e_comp	3.3	0.9	1	5	15915
Big five: agreeableness						
	big5_a_comp	3.5	0.7	1	5	15915
Big five: conscientiousness						
	big5_c_comp	3.2	0.8	1	5	15915
Big five: neuroticism						
	big5_n_comp	2.9	0.9	1	5	15915
Big five: openness						
	big5_o_comp	3.3	0.9	1	5	15915

Share of cases with imputed missing values: 1.4%

Item descriptives	Indicators	Mean	Std. dev.	Min.	Max.	Valid obs.	
Big five: extraversion							
	bigfive1	3.1	1.1	1	5	15890	*
	bigfive6	3.6	1.0	1	5	15851	
Big five: agreeableness							
	bigfive2	3.2	1.1	1	5	15879	
	bigfive7	3.3	1.0	1	5	15854	*
	bigfive11	3.8	1.0	1	5	15838	
Big five: conscientiousness							
	bigfive3	2.8	1.1	1	5	15863	*
	bigfive8	3.6	0.9	1	5	15854	
Big five: neuroticism							
	bigfive4	2.8	1.1	1	5	15875	*
	bigfive9	3.0	1.1	1	5	15869	
Big five: openness							
	bigfive5	3.0	1.4	1	5	15875	*
	bigfive10	3.7	1.1	1	5	15864	

* Item category order reversed for composit calculation (see Rammstedt et al., 2007)

List of scales (wave 0)

Locus of control

Baseline survey sample

Composite descriptives	Variable name	Mean	Std. dev.	Min.	Max.	Obs.
Internal locus of control						
	loci_comp	4.0	0.7	1	5	15833
External locus of control						
	loce_comp	2.5	0.9	1	5	15833

Share of cases with imputed missing values: 0.6%

Item descriptives	Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
Internal locus of control						
	loci1	3.9	0.9	1	5	15811
	loci2	4.2	0.8	1	5	15812
External locus of control						
	loce1	2.3	1.1	1	5	15793
	loce2	2.6	1.1	1	5	15777

List of scales (wave 0)

Scale: Work-related extrinsic values

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	6673	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			96617
	Bayesian Information Criterion (BIC)			96686
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.668

Reliability and Dimensionality

Ordinal Cronbach's Alpha .655
(Cronbach's alpha = .560)
McDonald's Omega .658

Test of (one-)dimensionality (parallel analysis)
Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue *
factor 1	.96
factor 2	-.14
factor 3	-.20

* No component with an adjusted eigenvalue ≥ 1

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
vawe1	0.70	0.01	0.68	0.71
vawe2	0.62	0.01	0.60	0.63
vawe4	0.56	0.01	0.54	0.58

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
vawe1	3.2	0.7	1	4	16066
vawe2	3.7	0.6	1	4	16064
vawe4	2.9	0.9	1	4	16065

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
vawe1	1.80	-5.36	-2.46	1.06
vawe2	1.42	-5.41	-3.92	-1.02
vawe4	1.19	-3.30	-0.98	1.39

List of scales (wave 0)

Scale: Work-related extrinsic values (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	273	18	.000	237	9	.000	19	9	.026

Tests of measurement invariance across ...

	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	12	4	.016	7	2	.033	1	2	.629
Strong invariance (plus equal intercepts)	86	4	.000	21	2	.000	0	2	.815
Strict invariance (plus equal error variances)	90	4	.000	6	2	.050	6	2	.043

Configural factor similarity across ...

Tucker's congruence coefficient	Survey languages	TCC	Survey settings	TCC	Survey modes	TCC
	German vs. French	.997	classroom vs. unproctored	.999	web vs. PAP	.997
	French vs. Italian	.988				
	Italian vs. German	.997				

Factor score equivalence: group specific vs. invariant models for ...

Coefficient of determination	Survey languages	CD	Survey settings	CD	Survey modes	CD
	German	1.000	classroom	1.000	web	1.000
	French	.994	unproctored	.995	PAP	.988
	Italian	.977				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
vawe_fs	0.0	0.7	-2.8	1.2	16084

Share of cases with imputed missing values: 0.3%

(Equivalence of scores from robust MLMV: CD = .996)

(Equivalence of Scores from Two-Step-Approach: CD = .975)

List of scales (wave 0)

Scale: Work-related intrinsic values

Baseline survey sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	14560	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			80533
	Bayesian Information Criterion (BIC)			80602
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker-Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.818

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.789
(Cronbach's alpha = .705)	
McDonald's Omega	.793

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	1.52
factor 2	-.11
factor 3	-.18

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
vawi1	0.72	0.01	0.71	0.73
vawi2	0.85	0.01	0.84	0.86
vawi5	0.67	0.01	0.66	0.68

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
vawi1	3.2	0.7	1	4	16078
vawi2	3.5	0.6	1	4	16071
vawi5	3.5	0.6	1	4	16065

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
vawi1	1.83	-5.30	-2.78	0.95
vawi2	3.18	-8.88	-6.16	-0.70
vawi5	1.64	-5.46	-3.70	-0.35

List of scales (wave 0)

Scale: Work-related intrinsic values (continued)

Baseline survey sample

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	376	18	.000	413	9	.000	32	9	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	2	4	.727	5	2	.075	24	2	.000
Strong invariance (plus equal intercepts)	179	4	.000	109	2	.000	1	2	.760
Strict invariance (plus equal error variances)	81	4	.000	3	2	.236	5	2	.070

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	1.000	classroom vs. unproctored	.999	web vs. PAP	.985
	French vs. Italian	1.000				
	Italian vs. German	1.000				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	1.000	web	.999
	French	1.000	unproctored	.999	PAP	.962
	Italian	1.000				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
vawi_fs	0.0	0.8	-3.0	1.1	16086

Share of cases with imputed missing values: 0.2%

(Equivalence of scores from robust MLMV: CD = .993)

(Equivalence of Scores from Two-Step-Approach: CD = .964)

List of scales (wave 0)

Family values

Baseline survey sample

Composite descriptives Variable name	Mean	Std. dev.	Min.	Max.	Obs.
vafa_comp	3.1	0.8	1	4	16075
Share of cases with imputed missing values:	0.2%				

Item descriptives Indicators	Mean	Std. dev.	Min.	Max.	Valid obs.
vafa1	3.3	0.8	1	4	16064
vafa2	3.0	0.9	1	4	16051

List of scales (wave 0)

Scale: Positive attitude towards life

AES Extension Survey

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	1110	5	.000
	Baseline vs. saturated	13955	10	.000
2)	**Root mean squared error (RMSEA)**			.208
	90% Confidence interval: lower bound			.198
	90% Confidence interval: upper bound			.218
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			57850
	Bayesian Information Criterion (BIC)			57948
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.921
	Tucker-Lewis Index (TLI)			.841
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.050
	Coefficient of determination (CD)			.893

Reliability and Dimensionality

Ordinal Cronbach's Alpha .880

(Cronbach's alpha = .844)

McDonald's Omega .881

Test of (one-)dimensionality (parallel analysis)

Criterion: Retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
factor 1	2.91
factor 2	.18
factor 3	-.03
factor 4	-.13
factor 5	-.11

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
posl1	0.72	0.01	0.70	0.74
posl2	0.84	0.01	0.83	0.85
posl3	0.78	0.01	0.76	0.79
posl5	0.67	0.01	0.65	0.69
posl6	0.85	0.01	0.84	0.86

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
posl1	5.0	0.9	1	6	5106
posl2	5.4	0.9	1	6	5107
posl3	4.8	1.0	1	6	5106
posl5	4.6	1.1	1	6	5108
posl6	5.0	1.1	1	6	5103

List of scales (wave 0)

Tests and Indices of Factorial Invariance across ...

Equality of the variance-covariance matrices across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
	933	40	.000	/		/	146	20	.000

Tests of measurement invariance across ...	Survey languages			Survey settings			Survey modes		
	chi2	df	p > chi2	chi2	df	p > chi2	chi2	df	p > chi2
Metric invariance (equal factor loadings)	9	8	.385	/		/	17	4	.002
Strong invariance (plus equal intercepts)	311	8	.000	/		/	7	4	.113
Strict invariance (plus equal error variances)	282	8	.000	/		/	20	4	.001

Configural factor similarity across ...	Survey languages		Survey settings		Survey modes	
Tucker's congruence coefficient		TCC		TCC		TCC
	German vs. French	.999	classroom vs. unproctored	/	web vs. PAP	.998
	French vs. Italian	.998				
	Italian vs. German	1.000				

Factor score equivalence: group specific vs. invariant models for ...	Survey languages		Survey settings		Survey modes	
Coefficient of determination		CD		CD		CD
	German	1.000	classroom	/	web	1.000
	French	1.000	unproctored		PAP	.999
	Italian	.999				

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
posl_fs	0.0	0.6	-3.0	0.7	5114

Share of cases with imputed missing values: 0.5%
(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Reality-based learning

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	129	2	.000
	Baseline vs. saturated	14527	6	.000

2)	**Root mean squared error (RMSEA)**		.076
	90% Confidence interval:	lower bound	.065
	90% Confidence interval:	upper bound	.087
	Probability RMSEA <= 0.05		.000
3)	**Akaike's Information Criterion (AIC)**		145766
	Bayesian Information Criterion (BIC)		145853
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.991
	Tucker–Lewis Index (TLI)		.974
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.016
	Coefficient of determination (CD)		.832

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.807
(Cronbach's alpha = .779)	
McDonald's Omega	.811

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.94
Factor 2	-.04
Factor 3	-.11
Factor 4	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
realref1	0.61	.007	0.60	0.63
realref2	0.65	.007	0.64	0.66
realref3	0.80	.005	0.79	0.81
realref4	0.80	.005	0.79	0.81

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
realref1	3.8	1.5	1	6	11042
realref2	3.9	1.4	1	6	10995
realref3	3.7	1.5	1	6	10984
realref4	4.1	1.5	1	6	11035

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	388	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	210	6	.000
Strong invariance (plus equal intercepts)	116	6	.000
Strict invariance (plus equal error variances)	78	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.983
French vs. Italian language version	.993
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.999
Language: French	.989
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
realref_fs	0.0	0.8	-2.1	1.6	11063

Share of cases with imputed missing values: 1.1%

(Equivalence of scores from robust MLMV: CD = .998)

List of scales (wave 0)

Scale: Discovery / exploratory learning

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	132	2	.000
	Baseline vs. saturated	19790	6	.000
2)	**Root mean squared error (RMSEA)**			.076
	90% Confidence interval: lower bound			.066
	90% Confidence interval: upper bound			.088
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			143687
	Bayesian Information Criterion (BIC)			143775
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.993
	Tucker–Lewis Index (TLI)			.980
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.013
	Coefficient of determination (CD)			.867

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.858
(Cronbach's alpha = .836)	
McDonald's Omega	.859

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.30
Factor 2	-.06
Factor 3	-.09
Factor 4	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
disclearn1	0.73	.005	0.72	0.74
disclearn2	0.84	.004	0.83	0.85
disclearn3	0.81	.004	0.80	0.82
disclearn4	0.72	.005	0.71	0.74

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
disclearn1	3.5	1.6	1	6	11049
disclearn2	3.5	1.5	1	6	10986
disclearn3	3.6	1.5	1	6	11002
disclearn4	3.7	1.5	1	6	11006

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	712	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	83	6	.000
Strong invariance (plus equal intercepts)	126	6	.000
Strict invariance (plus equal error variances)	190	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.985
French vs. Italian language version	.992
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.993
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
disclearn_fs	0.0	1.1	-2.3	2.1	11067

Share of cases with imputed missing values: 1.1%
(Equivalence of scores from robust MLMV: CD = .998)

List of scales (wave 0)

Scale: Social learning

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	5090	9	.000
	Baseline vs. saturated	36459	15	.000
2)	**Root mean squared error (RMSEA)**			.226
	90% Confidence interval: lower bound			.221
	90% Confidence interval: upper bound			.231
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			211536
	Bayesian Information Criterion (BIC)			211668
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.861
	Tucker–Lewis Index (TLI)			.768
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.096
	Coefficient of determination (CD)			.912

Reliability and Dimensionality

Ordinal Cronbach's Alpha .869

(Cronbach's alpha = .849)

McDonald's Omega .865

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.20
Factor 2	.48
Factor 3	-.06
Factor 4	-.08
Factor 5	-.09
Factor 6	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
comlearn1	0.54	.007	0.52	0.55
comlearn2	0.51	.008	0.50	0.53
comlearn3	0.62	.006	0.61	0.64
soclearn1	0.83	.004	0.83	0.84
soclearn2	0.88	.003	0.87	0.89
soclearn3	0.87	.003	0.87	0.88

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
comlearn1	3.8	1.5	1	6	11035
comlearn2	3.5	1.5	1	6	11009
comlearn3	3.7	1.5	1	6	10993
soclearn1	4.0	1.6	1	6	11039
soclearn2	4.3	1.5	1	6	11004
soclearn3	4.2	1.5	1	6	10990

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	580	54	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	55	10	.000
Strong invariance (plus equal intercepts)	202	10	.000
Strict invariance (plus equal error variances)	155	10	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.997
Italian vs. German language version	.997

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.998

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
soccomlearn_fs	0.0	0.8	-1.9	1.2	11065

Share of cases with imputed missing values: 1.2%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Social learning: social arrangement

Maths sample-split

Model and Fit Statistics

1) **Likelihood-ratio tests**

	chi2	df	p > chi2
Model vs. saturated	0	0	
Baseline vs. saturated	21585	3	.000

2) **Root mean squared error (RMSEA)** .000

90% Confidence interval:	lower bound	.000
90% Confidence interval:	upper bound	.000
Probability RMSEA <= 0.05		1.000

3) **Akaike's Information Criterion (AIC)** 100479

Bayesian Information Criterion (BIC) 100545

4) **Baseline comparison**

Comparative Fit Index (CFI) 1.000

Tucker–Lewis Index (TLI) 1.000

5) **Size of residuals**

Stand. root mean squared residual (SRMR) .000

Coefficient of determination (CD) .914

Reliability and Dimensionality

Ordinal Cronbach's Alpha .904

(Cronbach's alpha = .882)

McDonald's Omega .905

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.16
Factor 2	-.07
Factor 3	-.11

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
soclearn1	0.85	.003	0.84	0.86
soclearn2	0.92	.003	0.92	0.93
soclearn3	0.84	.004	0.84	0.85

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
soclearn1	4.0	1.6	1	6	11039
soclearn2	4.3	1.5	1	6	11004
soclearn3	4.2	1.5	1	6	10990

List of scales (wave 0)

Scale: Social learning: social arrangement (continued) — Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	142	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	25	4	.000
Strong invariance (plus equal intercepts)	54	4	.000
Strict invariance (plus equal error variances)	21	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.999
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.999

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
soclearn_fs	0.0	1.2	-2.9	1.7	11060

Share of cases with imputed missing values: 1.0%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Social learning: communication

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	9617	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			111136
	Bayesian Information Criterion (BIC)			111202
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.816

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.782
(Cronbach's alpha = .751)	
McDonald's Omega	.786

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.47
Factor 2	-.10
Factor 3	-.18

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
comlearn1	0.70	.007	0.69	0.72
comlearn2	0.66	.007	0.65	0.68
comlearn3	0.85	.007	0.84	0.87

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
comlearn1	3.8	1.5	1	6	11035
comlearn2	3.5	1.5	1	6	11009
comlearn3	3.7	1.5	1	6	10993

List of scales (wave 0)

Scale: Social learning: communication (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	261	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	9	4	.070
Strong invariance (plus equal intercepts)	53	4	.000
Strict invariance (plus equal error variances)	17	4	.002

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.999
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
comlearn_fs	0.0	0.9	-2.1	1.8	11062

Share of cases with imputed missing values: 1.0%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Instructivist learning

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	4517	20	.000
	Baseline vs. saturated	29913	28	.000

2)	**Root mean squared error (RMSEA)**		.143
	90% Confidence interval:	lower bound	.139
	90% Confidence interval:	upper bound	.146
	Probability RMSEA <= 0.05		.000
3)	**Akaike's Information Criterion (AIC)**		286311
	Bayesian Information Criterion (BIC)		286487
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.850
	Tucker–Lewis Index (TLI)		.789
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.066
	Coefficient of determination (CD)		.848

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.841
(Cronbach's alpha = .818)	
McDonald's Omega	.842

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.18
Factor 2	.36
Factor 3	.21
Factor 4	.05
Factor 5	-.10
Factor 6	-.14
Factor 7	-.14
Factor 8	-.20

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
instrlearn1	0.65	.006	0.64	0.67
instrlearn2	0.65	.007	0.63	0.66
instrlearn3	0.48	.008	0.47	0.50
instrlearn4	0.70	.006	0.69	0.71
replearn1	0.67	.006	0.66	0.68
replearn2	0.59	.007	0.58	0.61
replearn3	0.60	.007	0.59	0.62
replearn4	0.70	.006	0.69	0.71

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
instrlearn1	4.6	1.4	1	6	11031
instrlearn2	3.8	1.4	1	6	11001
instrlearn3	3.3	1.5	1	6	10993
instrlearn4	4.6	1.4	1	6	11052
replearn1	4.4	1.4	1	6	11041
replearn2	4.3	1.3	1	6	10990
replearn3	3.6	1.4	1	6	10991
replearn4	4.3	1.4	1	6	11010

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	4066	88	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	117	14	.000
Strong invariance (plus equal intercepts)	1511	14	.000
Strict invariance (plus equal error variances)	337	14	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.994
French vs. Italian language version	.996
Italian vs. German language version	.990

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.999
Language: French	.998
Language: Italian	.993

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
instreplearn_fs	0.0	0.8	-2.7	1.5	11069

Share of cases with imputed missing values: 1.3%

(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Scale: Instructivist learning: teachers instructions

Maths sample-split

Model and Fit Statistics

1) **Likelihood-ratio tests**

	chi2	df	p > chi2
Model vs. saturated	605	2	.000
Baseline vs. saturated	9077	6	.000

2) **Root mean squared error (RMSEA)** .165

90% Confidence interval:	lower bound	.154
90% Confidence interval:	upper bound	.176
Probability RMSEA <= 0.05		.000

3) **Akaike's Information Criterion (AIC)** 147556

Bayesian Information Criterion (BIC) 147643

4) **Baseline comparison**

Comparative Fit Index (CFI)	.934
Tucker–Lewis Index (TLI)	.801

5) **Size of residuals**

Stand. root mean squared residual (SRMR)	.045
Coefficient of determination (CD)	.741

Reliability and Dimensionality

Ordinal Cronbach's Alpha .723

(Cronbach's alpha = .683)

McDonald's Omega .727

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.48
Factor 2	.05
Factor 3	-.12
Factor 4	-.22

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
instrlearn1	0.66	.008	0.65	0.68
instrlearn2	0.68	.008	0.67	0.70
instrlearn3	0.49	.009	0.47	0.51
instrlearn4	0.69	.008	0.67	0.70

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
instrlearn1	4.6	1.4	1	6	11031
instrlearn2	3.8	1.4	1	6	11001
instrlearn3	3.3	1.5	1	6	10993
instrlearn4	4.6	1.4	1	6	11052

List of scales (wave 0)

Scale: Instructivist learning: teachers instructions (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	2118	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	49	6	.000
Strong invariance (plus equal intercepts)	466	6	.000
Strict invariance (plus equal error variances)	146	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.994
French vs. Italian language version	.975
Italian vs. German language version	.978

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.998
Language: French	.998
Language: Italian	.958

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
instrlearn_fs	0.0	0.8	-2.6	1.4	11064

Share of cases with imputed missing values: 1.1%

(Equivalence of scores from robust MLMV: CD = .989)

List of scales (wave 0)

Scale: Instructivist learning: repetitive practice

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	24	2	.000
	Baseline vs. saturated	9920	6	.000

2)	**Root mean squared error (RMSEA)**		.032
	90% Confidence interval:	lower bound	.021
	90% Confidence interval:	upper bound	.043
	Probability RMSEA <= 0.05		.996
3)	**Akaike's Information Criterion (AIC)**		145662
	Bayesian Information Criterion (BIC)		145750
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.998
	Tucker–Lewis Index (TLI)		.993
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.008
	Coefficient of determination (CD)		.774

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.745
(Cronbach's alpha = .713)	
McDonald's Omega	.751

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.58
Factor 2	-.08
Factor 3	-.10
Factor 4	-.16

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
replearn1	0.76	.007	0.75	0.78
replearn2	0.71	.007	0.70	0.72
replearn3	0.49	.009	0.48	0.51
replearn4	0.64	.007	0.63	0.66

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
replearn1	4.4	1.4	1	6	11041
replearn2	4.3	1.3	1	6	10990
replearn3	3.6	1.4	1	6	10991
replearn4	4.3	1.4	1	6	11010

List of scales (wave 0)

Scale: Instructivist learning: repetitive practice (continued)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1353	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	36	6	.000
Strong invariance (plus equal intercepts)	965	6	.000
Strict invariance (plus equal error variances)	209	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.996
French vs. Italian language version	.999
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.999
Language: French	.996
Language: Italian	.997

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
replearn_fs	0.0	0.9	-2.8	1.5	11067

Share of cases with imputed missing values: 1.1%

(Equivalence of scores from robust MLMV: CD = .997)

List of scales (wave 0)

Scale: System aspect

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	2443	9	.000
	Baseline vs. saturated	31459	15	.000
2)	**Root mean squared error (RMSEA)**			.157
	90% Confidence interval: lower bound			.152
	90% Confidence interval: upper bound			.162
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			185422
	Bayesian Information Criterion (BIC)			185553
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.923
	Tucker–Lewis Index (TLI)			.871
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.050
	Coefficient of determination (CD)			.879

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.878
(Cronbach's alpha = .854)	
McDonald's Omega	.878

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.21
Factor 2	.22
Factor 3	-.03
Factor 4	-.06
Factor 5	-.13
Factor 6	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
formasp1	0.71	.006	0.70	0.73
formasp2	0.72	.005	0.71	0.73
formasp3	0.75	.005	0.74	0.76
systasp1	0.74	.005	0.73	0.75
systasp2	0.76	.005	0.75	0.77
systasp3	0.75	.005	0.74	0.76

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
formasp1	4.3	1.3	1	6	10946
formasp2	4.1	1.3	1	6	10932
formasp3	4.4	1.2	1	6	10965
systasp1	5.0	1.2	1	6	10967
systasp2	4.7	1.2	1	6	10925
systasp3	4.7	1.2	1	6	10975

List of scales (wave 0)

Scale: System aspect (continued)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	478	54	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	64	10	.000
Strong invariance (plus equal intercepts)	171	10	.000
Strict invariance (plus equal error variances)	45	10	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.998
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.998

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
sysformasp_fs	0.0	0.8	-3.2	1.3	11006

Share of cases with imputed missing values: 1.3%
(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: System aspect: logical thinking

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	12550	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			92905
	Bayesian Information Criterion (BIC)			92970
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.833

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.832
(Cronbach's alpha = .792)	
McDonald's Omega	.832

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.70
Factor 2	-.13
Factor 3	-.14

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
systasp1	0.76	.006	0.75	0.78
systasp2	0.81	.005	0.79	0.82
systasp3	0.80	.005	0.79	0.81

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
systasp1	5.0	1.2	1	6	10967
systasp2	4.7	1.2	1	6	10925
systasp3	4.7	1.2	1	6	10975

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	210	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	35	4	.000
Strong invariance (plus equal intercepts)	84	4	.000
Strict invariance (plus equal error variances)	13	4	.012

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.997
French vs. Italian language version	1.000
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.997
Language: Italian	.995

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
systasp_fs	0.0	0.8	-3.1	1.0	11004

Share of cases with imputed missing values: 1.0%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: System aspect: formalism

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	11712	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			97123
	Bayesian Information Criterion (BIC)			97189
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.822

Reliability and Dimensionality

Ordinal Cronbach's Alpha .821

(Cronbach's alpha = .791)

McDonald's Omega .821

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.65
Factor 2	-.14
Factor 3	-.14

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
formasp1	0.78	.006	0.77	0.79
formasp2	0.79	.006	0.78	0.80
formasp3	0.77	.006	0.76	0.78

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
formasp1	4.3	1.3	1	6	10946
formasp2	4.1	1.3	1	6	10932
formasp3	4.4	1.2	1	6	10965

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	193	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	11	4	.025
Strong invariance (plus equal intercepts)	83	4	.000
Strict invariance (plus equal error variances)	14	4	.008

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.994
Italian vs. German language version	.993

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.985

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
formasp_fs	0.0	0.9	-2.7	1.5	10992

Share of cases with imputed missing values: 0.9%
(Equivalence of scores from robust MLMV: CD = 1.00)

List of scales (wave 0)

Scale: Scheme aspect

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	12713	3	.000

2)	**Root mean squared error (RMSEA)**		.000
	90% Confidence interval:	lower bound	.000
	90% Confidence interval:	upper bound	.000
	Probability RMSEA <= 0.05		1.000
3)	**Akaike's Information Criterion (AIC)**		100471
	Bayesian Information Criterion (BIC)		100537
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		1.000
	Tucker–Lewis Index (TLI)		1.000
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.000
	Coefficient of determination (CD)		.843

Reliability and Dimensionality

Ordinal Cronbach's Alpha .832
(Cronbach's alpha = .806)
McDonald's Omega .833

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.72
Factor 2	-.11
Factor 3	-.16

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
schemasp1	0.76	.006	0.75	0.77
schemasp2	0.76	.006	0.75	0.77
schemasp3	0.85	.005	0.84	0.86

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
schemasp1	3.9	1.4	1	6	10967
schemasp2	4.0	1.3	1	6	10926
schemasp3	3.7	1.4	1	6	10927

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	313	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	8	4	.092
Strong invariance (plus equal intercepts)	98	4	.000
Strict invariance (plus equal error variances)	25	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	1.000
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.998

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
schemasp_fs	0.0	0.9	-2.4	1.8	10990

Share of cases with imputed missing values: 0.9%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Application aspect

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	316	2	.000
	Baseline vs. saturated	20302	6	.000
2)	**Root mean squared error (RMSEA)**			.119
	90% Confidence interval: lower bound			.109
	90% Confidence interval: upper bound			.131
	Probability RMSEA <= 0.05			.000
3)	**Akaike's Information Criterion (AIC)**			129471
	Bayesian Information Criterion (BIC)			129559
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			.985
	Tucker–Lewis Index (TLI)			.954
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.021
	Coefficient of determination (CD)			.866

Reliability and Dimensionality

Ordinal Cronbach's Alpha .863

(Cronbach's alpha = .839)

McDonald's Omega .864

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.33
Factor 2	-.03
Factor 3	-.11
Factor 4	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
applyasp1	0.80	.005	0.79	0.81
applyasp2	0.79	.005	0.78	0.80
applyasp3	0.73	.005	0.72	0.74
applyasp4	0.81	.005	0.80	0.82

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
applyasp1	4.2	1.3	1	6	10982
applyasp2	4.6	1.3	1	6	10933
applyasp3	3.9	1.4	1	6	10958
applyasp4	4.3	1.3	1	6	10924

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	498	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	70	6	.000
Strong invariance (plus equal intercepts)	151	6	.000
Strict invariance (plus equal error variances)	53	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.997
French vs. Italian language version	.992
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.999

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
applyasp_fs	0.0	1.0	-3.0	1.6	11007

Share of cases with imputed missing values: 1.1%

(Equivalence of scores from robust MLMV: CD = .999)

List of scales (wave 0)

Scale: Teacher: cognitive activation
Maths sample-split

Model and Fit Statistics

1) **Likelihood-ratio tests**

	chi2	df	p > chi2
Model vs. saturated	5636	20	.000
Baseline vs. saturated	38613	28	.000

2) **Root mean squared error (RMSEA)** .164

90% Confidence interval:	lower bound	.160
90% Confidence interval:	upper bound	.167
Probability RMSEA <= 0.05		.000

3) **Akaike's Information Criterion (AIC)** 176245
Bayesian Information Criterion (BIC) 176419

4) **Baseline comparison**

Comparative Fit Index (CFI)	.854
Tucker–Lewis Index (TLI)	.796

5) **Size of residuals**

Stand. root mean squared residual (SRMR)	.078
Coefficient of determination (CD)	.894

Reliability and Dimensionality

Ordinal Cronbach's Alpha .873
(Cronbach's alpha = .844)
McDonald's Omega .872

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.74
Factor 2	.52
Factor 3	.15
Factor 4	-.03
Factor 5	-.07
Factor 6	-.13
Factor 7	-.14
Factor 8	-.14

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
cogself1	0.83	.004	0.82	0.83
cogself2	0.50	.008	0.48	0.51
cogself3	0.56	.007	0.54	0.57
cogself4	0.75	.005	0.74	0.76
cogself5	0.82	.004	0.81	0.83
cogself6	0.66	.006	0.64	0.67
cogself7	0.62	.007	0.61	0.63
cogself8	0.67	.006	0.66	0.68

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
cogself1	2.8	0.9	1	4	10443
cogself2	2.6	0.8	1	4	10290
cogself3	2.7	0.9	1	4	10324
cogself4	2.9	0.8	1	4	10423
cogself5	2.8	0.9	1	4	10428
cogself6	2.9	0.8	1	4	10432
cogself7	2.7	0.8	1	4	10271
cogself8	2.7	0.8	1	4	10278

Parameters of Generalized Structural Equation Model (Ordinal Logit Link)

Indicators	Coef.	Cut1	Cut2	Cut3
cogself1	2.85	-4.53	-1.62	2.57
cogself2	1.13	-2.48	-0.26	2.42
cogself3	1.29	-2.66	-0.59	2.07
cogself4	2.17	-3.98	-1.53	1.87
cogself5	2.75	-4.35	-1.27	2.61
cogself6	1.67	-3.68	-1.26	1.58
cogself7	1.56	-3.22	-0.69	2.66
cogself8	1.77	-3.44	-0.88	2.53

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	943	88	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	46	14	.000
Strong invariance (plus equal intercepts)	495	14	.000
Strict invariance (plus equal error variances)	321	14	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.998
Italian vs. German language version	.996

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.999

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cogself_fs	0.0	0.9	-2.8	2.3	10496

Share of cases with imputed missing values: 3.2%
(Equivalence of scores from robust MLMV: CD = .998)
(Equivalence of scores from two-step approach: CD = .983)

List of scales (wave 0)

Scale: Cogn. activation: finding solutions & arguing

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	332	2	.000
	Baseline vs. saturated	19997	6	.000

2)	**Root mean squared error (RMSEA)**		.125
	90% Confidence interval:	lower bound	.114
	90% Confidence interval:	upper bound	.137
	Probability RMSEA <= 0.05		.000
3)	**Akaike's Information Criterion (AIC)**		85451
	Bayesian Information Criterion (BIC)		85538
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.984
	Tucker–Lewis Index (TLI)		.951
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.023
	Coefficient of determination (CD)		.878

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.864
(Cronbach's alpha = .825)	
McDonald's Omega	.865

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.37
Factor 2	-.02
Factor 3	-.08
Factor 4	-.15

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
cogself1	0.83	.004	0.83	0.84
cogself4	0.75	.005	0.74	0.76
cogself5	0.86	.004	0.85	0.86
cogself6	0.69	.006	0.68	0.71

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
cogself1	2.8	0.9	1	4	10443
cogself4	2.9	0.8	1	4	10423
cogself5	2.8	0.9	1	4	10428
cogself6	2.9	0.8	1	4	10432

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
cogself1	2.72	-4.62	-1.74	2.31
cogself4	2.19	-4.18	-1.71	1.71
cogself5	3.03	-4.91	-1.50	2.61
cogself6	1.91	-4.07	-1.49	1.53

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	351	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	24	6	.000
Strong invariance (plus equal intercepts)	110	6	.000
Strict invariance (plus equal error variances)	105	6	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.997
Italian vs. German language version	.997

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.995

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cogselfa_fs	-0.1	0.9	-2.4	1.7	10467

Share of cases with imputed missing values: 0.9%
(Equivalence of scores from robust MLMV: CD = .999)
(Equivalence of scores from two-step approach: CD = .985)

List of scales (wave 0)

Scale: Cogn. activation: strategies & learning from mistakes — Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	1037	2	.000
	Baseline vs. saturated	12679	6	.000

2)	**Root mean squared error (RMSEA)**	.224
	90% Confidence interval: lower bound	.212
	90% Confidence interval: upper bound	.235
	Probability RMSEA <= 0.05	.000
3)	**Akaike's Information Criterion (AIC)**	90475
	Bayesian Information Criterion (BIC)	90562
4)	**Baseline comparison**	
	Comparative Fit Index (CFI)	.918
	Tucker–Lewis Index (TLI)	.755
5)	**Size of residuals**	
	Stand. root mean squared residual (SRMR)	.061
	Coefficient of determination (CD)	.816

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.788
(Cronbach's alpha = .743)	
McDonald's Omega	.787

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.84
Factor 2	.12
Factor 3	-.18
Factor 4	-.17

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
cogself2	0.60	.008	0.59	0.62
cogself3	0.58	.008	0.56	0.59
cogself7	0.76	.006	0.75	0.78
cogself8	0.81	.006	0.80	0.82

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
cogself2	2.6	0.8	1	4	10290
cogself3	2.7	0.9	1	4	10324
cogself7	2.7	0.8	1	4	10271
cogself8	2.7	0.8	1	4	10278

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
cogself2	1.45	-2.80	-0.35	2.62
cogself3	1.36	-2.79	-0.64	2.10
cogself7	2.13	-3.95	-0.89	3.12
cogself8	2.37	-4.28	-1.14	3.01

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	402	28	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	70	6	.000
Strong invariance (plus equal intercepts)	151	6	.000
Strict invariance (plus equal error variances)	124	6	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.962
Italian vs. German language version	.975

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.999
Language: Italian	.936

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
cogselfb_fs	0.0	0.9	-2.4	2.1	10334

Share of cases with imputed missing values: 0.9%

(Equivalence of scores from robust MLMV: CD = .996)

(Equivalence of scores from two-step approach: CD = .985)

List of scales (wave 0)

Scale: Teacher: classroom management

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	16993	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			63509
	Bayesian Information Criterion (BIC)			63574
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.892

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.882
(Cronbach's alpha = .842)	
McDonald's Omega	.883

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.02
Factor 2	-.09
Factor 3	-.12

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
classman1	0.79	.005	0.78	0.80
classman2	0.85	.004	0.84	0.85
classman3	0.90	.004	0.89	0.90

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
classman1	2.4	0.9	1	4	10313
classman2	2.4	0.9	1	4	10295
classman3	2.3	0.9	1	4	10272

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
classman1	2.48	-3.02	0.54	3.83
classman2	3.05	-3.19	0.28	4.06
classman3	3.96	-3.53	0.98	5.59

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	267	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	6	4	.169
Strong invariance (plus equal intercepts)	58	4	.000
Strict invariance (plus equal error variances)	13	4	.010

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	.999
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	.999

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
classman_fs	0.0	0.9	-1.7	2.0	10343

Share of cases with imputed missing values: 0.9%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of scores from two-step approach: CD = .992)

List of scales (wave 0)

Scale: Teacher: individual learning support

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	121	5	.000
	Baseline vs. saturated	42736	10	.000

2)	**Root mean squared error (RMSEA)**		.047
	90% Confidence interval:	lower bound	.040
	90% Confidence interval:	upper bound	.055
	Probability RMSEA <= 0.05		.730
3)	**Akaike's Information Criterion (AIC)**		94824
	Bayesian Information Criterion (BIC)		94932
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		.997
	Tucker–Lewis Index (TLI)		.995
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.007
	Coefficient of determination (CD)		.936

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.935
(Cronbach's alpha = .907)	
McDonald's Omega	.935

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	3.63
Factor 2	-.04
Factor 3	-.06
Factor 4	-.05
Factor 5	-.06

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
indsup1	0.86	.003	0.85	0.86
indsup2	0.89	.003	0.88	0.89
indsup3	0.87	.003	0.87	0.88
indsup4	0.87	.003	0.86	0.87
indsup5	0.82	.004	0.81	0.83

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
indsup1	2.7	0.9	1	4	10434
indsup2	3.0	0.9	1	4	10436
indsup3	2.8	0.9	1	4	10464
indsup4	2.8	0.9	1	4	10439
indsup5	2.9	0.9	1	4	10423

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
indsup1	3.14	-4.44	-1.26	2.84
indsup2	3.72	-5.69	-2.62	1.91
indsup3	3.43	-4.89	-1.86	2.64
indsup4	3.29	-4.42	-1.53	2.12
indsup5	2.74	-4.43	-1.76	2.14

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	515	40	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	35	8	.000
Strong invariance (plus equal intercepts)	196	8	.000
Strict invariance (plus equal error variances)	57	8	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.999
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	1.000
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
indsup_fs	0.0	0.9	-2.2	1.6	10486

Share of cases with imputed missing values: 1.0%
(Equivalence of scores from robust MLMV: CD = 1.00)
(Equivalence of scores from two-step approach: CD = .981)

List of scales (wave 0)

Scale: Teacher: instruction quality

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	9348	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			71991
	Bayesian Information Criterion (BIC)			72056
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.829

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.765
(Cronbach's alpha = .712)	
McDonald's Omega	.780

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.47
Factor 2	-.08
Factor 3	-.18

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
instqual1	0.80	.007	0.79	0.82
instqual2	0.85	.007	0.84	0.87
instqual3	0.53	.008	0.51	0.54

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
instqual1	2.8	0.9	1	4	10426
instqual2	2.8	0.8	1	4	10285
instqual3	2.6	0.9	1	4	10266

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
instqual1	2.52	-3.80	-1.28	1.99
instqual2	3.09	-4.94	-1.54	3.53
instqual3	1.15	-2.11	-0.25	2.18

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	432	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	31	4	.000
Strong invariance (plus equal intercepts)	310	4	.000
Strict invariance (plus equal error variances)	21	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.996
French vs. Italian language version	.999
Italian vs. German language version	.999

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.998
Language: Italian	.996

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
instqual_fs	0.0	0.9	-2.0	1.7	10473

Share of cases with imputed missing values: 2.6%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of scores from two-step approach: CD = .988)

List of scales (wave 0)

Scale: Situational interest

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	11000	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			76347
	Bayesian Information Criterion (BIC)			76413
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.834

Reliability and Dimensionality

Ordinal Cronbach's Alpha .806
(Cronbach's alpha = .757)
McDonald's Omega .810

Test of (one-)dimensionality (parallel analysis)
Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.60
Factor 2	-.10
Factor 3	-.17

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
intsit1	0.75	.006	0.73	0.76
intsit2	0.68	.007	0.67	0.70
intsit3	0.86	.006	0.85	0.87

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
intsit1	2.6	0.9	1	4	10891
intsit2	2.3	0.9	1	4	10836
intsit3	2.4	0.9	1	4	10897

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
intsit1	2.09	-3.06	-0.39	3.19
intsit2	1.82	-1.86	0.46	3.11
intsit3	3.24	-2.76	0.54	4.35

List of scales (wave 0)

Scale: Situational interest (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	801	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	282	4	.000
Strong invariance (plus equal intercepts)	61	4	.000
Strict invariance (plus equal error variances)	251	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.974
French vs. Italian language version	.999
Italian vs. German language version	.983

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.998
Language: French	.971
Language: Italian	.995

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
intsit_fs	0.0	0.9	-1.7	2.0	10926

Share of cases with imputed missing values: 1.2%

(Equivalence of scores from robust MLMV: CD = .996)

(Equivalence of scores from two-step approach: CD = .988)

List of scales (wave 0)

Scale: Perceived autonomy support

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	10030	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			72281
	Bayesian Information Criterion (BIC)			72346
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.809

Reliability and Dimensionality

Ordinal Cronbach's Alpha .799

(Cronbach's alpha = .748)

McDonald's Omega .800

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.55
Factor 2	-.13
Factor 3	-.17

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
persuppauto1	0.74	.006	0.73	0.76
persuppauto2	0.82	.006	0.81	0.83
persuppauto3	0.70	.007	0.69	0.72

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
persuppauto1	2.7	0.9	1	4	10665
persuppauto2	2.9	0.9	1	4	10627
persuppauto3	3.0	0.8	1	4	10655

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
persuppauto1	2.02	-3.46	-0.78	2.16
persuppauto2	2.67	-4.43	-1.76	2.02
persuppauto3	1.88	-4.13	-1.81	1.12

List of scales (wave 0)

Scale: Perceived autonomy support (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	229	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	34	4	.000
Strong invariance (plus equal intercepts)	142	4	.000
Strict invariance (plus equal error variances)	28	4	.000

Configural factor similarity	
Tucker's Congruence Coefficient	TCC
German vs. French language version	.996
French vs. Italian language version	.994
Italian vs. German language version	.998

Factor score equivalence: group specific vs. invariant models	
Coefficient of determination	CD
Language: German	1.000
Language: French	.993
Language: Italian	.993

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
persuppauto_fs	0.0	0.9	-2.2	1.5	10674

Share of cases with imputed missing values: 0.5%

(Equivalence of scores from robust MLMV: CD = .999)

(Equivalence of scores from two-step approach: CD = .987)

List of scales (wave 0)

Scale: Perceived competence support

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	19504	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			61112
	Bayesian Information Criterion (BIC)			61178
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.951

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.888
(Cronbach's alpha = .842)	
McDonald's Omega	.892

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	2.09
Factor 2	-.03
Factor 3	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
persuppcomp1	0.97	.003	0.96	0.98
persuppcomp2	0.77	.005	0.77	0.78
persuppcomp3	0.82	.004	0.81	0.83

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
persuppcomp1	2.9	0.8	1	4	10639
persuppcomp2	2.7	0.9	1	4	10639
persuppcomp3	3.0	0.8	1	4	10645

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
persuppcomp1	4.74	-7.76	-3.07	3.05
persuppcomp2	2.29	-3.63	-0.99	2.34
persuppcomp3	2.73	-5.44	-2.51	1.35

List of scales (wave 0)

Scale: Perceived competence support (continued) Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	281	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	61	4	.000
Strong invariance (plus equal intercepts)	124	4	.000
Strict invariance (plus equal error variances)	43	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.998
French vs. Italian language version	.998
Italian vs. German language version	.997

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.998
Language: Italian	.982

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
persuppcomp_fs	0.0	0.9	-2.2	1.5	10665

Share of cases with imputed missing values: 0.5%

(Equivalence of scores from robust MLMV: CD = .994)

(Equivalence of scores from two-step approach: CD = .953)

List of scales (wave 0)

Scale: Perceived social relatedness

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	15653	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			69393
	Bayesian Information Criterion (BIC)			69459
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.886

Reliability and Dimensionality

Ordinal Cronbach's Alpha .858

(Cronbach's alpha = .814)

McDonald's Omega .862

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.90
Factor 2	-.08
Factor 3	-.13

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
persocincl1	0.89	.004	0.88	0.89
persocincl2	0.70	.006	0.69	0.71
persocincl3	0.87	.004	0.86	0.88

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
persocincl1	2.7	0.9	1	4	10635
persocincl2	2.7	0.9	1	4	10640
persocincl3	2.4	0.9	1	4	10632

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
persocincl1	3.65	-4.81	-1.00	3.37
persocincl2	1.82	-3.18	-0.78	2.11
persocincl3	3.34	-2.89	0.28	4.36

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	1205	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	74	4	.000
Strong invariance (plus equal intercepts)	745	4	.000
Strict invariance (plus equal error variances)	216	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.993
French vs. Italian language version	.993
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	1.000
Language: French	.992
Language: Italian	1.000

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
persocincl_fs	0.0	0.9	-1.9	1.8	10684

Share of cases with imputed missing values: 0.9%
(Equivalence of scores from robust MLMV: CD = .996)
(Equivalence of scores from two-step approach: CD = .987)

List of scales (wave 0)

Scale: Classmates' appreciation of mathematics

Maths sample-split

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	19804	3	.000
2)	**Root mean squared error (RMSEA)**			.000
	90% Confidence interval: lower bound			.000
	90% Confidence interval: upper bound			.000
	Probability RMSEA <= 0.05			1.000
3)	**Akaike's Information Criterion (AIC)**			53455
	Bayesian Information Criterion (BIC)			53521
4)	**Baseline comparison**			
	Comparative Fit Index (CFI)			1.000
	Tucker–Lewis Index (TLI)			1.000
5)	**Size of residuals**			
	Stand. root mean squared residual (SRMR)			.000
	Coefficient of determination (CD)			.946

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.834
(Cronbach's alpha = .776)	
McDonald's Omega	.859

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.94
Factor 2	-.02
Factor 3	-.08

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf. interval]	
apprmath1	0.92	.004	0.92	0.93
apprmath2	0.96	.004	0.95	0.97
apprmath3	0.53	.007	0.51	0.54

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
apprmath1	2.0	0.7	1	4	10778
apprmath2	2.0	0.7	1	4	10775
apprmath3	2.7	0.8	1	4	10776

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
apprmath1	4.34	-2.78	3.80	8.49
apprmath2	4.83	-2.94	4.63	9.65
apprmath3	1.14	-2.82	-0.55	2.41

List of scales (wave 0)

Scale: Classmates' appreciation of mathematics (continued)

Maths sample-split

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	320	9	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	13	2	.001
Strong invariance (plus equal intercepts)	67	2	.000
Strict invariance (plus equal error variances)	5	2	.082

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	1.000
French vs. Italian language version	
Italian vs. German language version	

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.999
Language: French/ Italian	.991

*** Note:** Due to sparse tables for the italian version of the scale, equivalence tests failed to converge and were reestimated with collapsed italian and french versions.

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
apprmath_fs	0.0	0.9	-1.6	2.4	10784

Share of cases with imputed missing values: 0.1%

(Equivalence of scores from robust MLMV: CD = .997)

(Equivalence of scores from two-step approach: CD = .980)

List of scales (wave 0)

Scale: Absenteeism / truancy

Full AES sample

Model and Fit Statistics

		chi2	df	p > chi2
1)	**Likelihood-ratio tests**			
	Model vs. saturated	0	0	
	Baseline vs. saturated	30122	3	.000

2)	**Root mean squared error (RMSEA)**		.000
	90% Confidence interval:	lower bound	.000
	90% Confidence interval:	upper bound	.000
	Probability RMSEA <= 0.05		1.000
3)	**Akaike's Information Criterion (AIC)**		84033
	Bayesian Information Criterion (BIC)		84105
4)	**Baseline comparison**		
	Comparative Fit Index (CFI)		1.000
	Tucker–Lewis Index (TLI)		1.000
5)	**Size of residuals**		
	Stand. root mean squared residual (SRMR)		.000
	Coefficient of determination (CD)		.923

Reliability and Dimensionality

Ordinal Cronbach's Alpha	.819
(Cronbach's alpha = .648)	
McDonald's Omega	.837

Test of (one-)dimensionality (parallel analysis)

Criterion: retain factors with adj. eigenvalue > 0

	Adjusted eigenvalue
Factor 1	1.77
Factor 2	-.03
Factor 3	-.14

Standardized factor loadings

Indicators	Coef.	(SE)	[95% Conf.	interval]
truancy1	0.84	.004	0.83	0.85
truancy2 *	0.95	.004	0.94	0.96
truancy3 *	0.56	.005	0.55	0.57

*** Note:** Original item from TREE1 / PISA2000

Item descriptives

Indicators	Mean	Std. dev.	Min.	Max.	Valid Obs.
truancy1	1.1	0.4	1	4	22242
truancy2 *	1.2	0.5	1	4	22245
truancy3 *	1.5	0.8	1	4	22251

Parameters of generalized structural equation model (ordinal logit link)

Indicators	Coef.	Cut1	Cut2	Cut3
truancy1	3.27	4.85	7.51	8.62
truancy2	4.63	5.31	8.79	10.99
truancy3	1.16	0.54	2.44	3.49

List of scales (wave 0)

Tests and Indices of Factorial Invariance across Survey Languages

Equality of variance-covariance matrices	chi2	df	p > chi2
	2001	18	.000

Tests of measurement invariance	chi2	df	p > chi2
Metric invariance (equal factor loadings)	38	4	.000
Strong invariance (plus equal intercepts)	734	4	.000
Strict invariance (plus equal error variances)	680	4	.000

Configural factor similarity

Tucker's Congruence Coefficient	TCC
German vs. French language version	.999
French vs. Italian language version	.998
Italian vs. German language version	1.000

Factor score equivalence: group specific vs. invariant models

Coefficient of determination	CD
Language: German	.997
Language: French	.988
Language: Italian	.954

Factor score descriptives

Variable name	Mean	Std. dev.	Min.	Max.	Obs.
truancy_fs	0.0	0.7	-0.5	2.8	22254

Share of cases with imputed missing values: 0.1%
(Equivalence of scores from robust MLMV: CD = .995)
(Equivalence of scores from two-step approach: CD = .780)

List of scales (wave 0)

List of Sources

AEQ-M (short-version)

Pekrun, R. , Goetz, T. & Frenzel, A. C. (2005). Achievement emotions questionnaire-mathematics (AEQ-M). User's manual. Department of Psychology, University of Munich.

BHPS

Taylor, M. F., Brice, J., Buck, N., & Prentice-Lane, E. (2018). British Household Panel Survey User Documenation (Questionnaires Wave 13 & 15). Publication no. http://doi.org/10.5255/UKDA-SN-5151-2). University of Essex, Institute for Social and Economic Research.

Böhm-Kasper et al., 2000

Böhm-Kasper, O., Bos, W., Jaeckel, S. & Weishaupt, H. (2000). Skalenhandbuch zur Belastung von Schülern und Lehrern. Das Erfurter Belastungs-Inventar (EBI). Erfurt: Pädagogische Hochschule Erfurt.

COACTIV 2008

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M., & Tsai, Y.-M. (2008). Professionswissen von Lehrkräften, kognitiv aktivierender Mathematik-unterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente. Berlin: Max-Planck-Institut für Bildungsforschung.

Eder, 1995, 2007

Eder, F. (Ed.) (1995). Das Befinden von Kindern und Jugendlichen in der Schule. Innsbruck: Studienverlag.

Eder, F. (2007): Das Befinden von Kindern und Jugendlichen in der österreichischen Schule. Befragung 2005.

GESIS (short-version)

Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2012). Eine Kurzskala zur Messung von Kontrollüberzeugung: die Skala Internale-Externale-Kontrollüberzeugung-4 (IE-4) (GESIS-Working Papers, 2012/19). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-31209

Girnat, 2015, 2017

Girnat, B. (2015). Girnat, B. (2015). Konstruktivistische und instruktivistische Lehrmethoden aus Schülersicht – Entwicklung eines fragebogenbasierten Erhebungsinstrumentes. In F. Caluori, H. Linneweber-Lammerskitten, & C. Streit (Eds.), Beiträge zum Mathematikunterricht 2015 (Vol. 1, pp. 308 – 311). Münster: WTM.

Girnat, B. (2017). Gender Differences Concerning Pupils' Beliefs on Teaching Methods and Mathematical Worldviews at Lower Secondary Schools. In: C. Andrà, D. Brunetto, E. Levenson und P. Liljedahl (Eds.): Teaching and Learning in Maths Classrooms: Emerging Themes in Affect-related Research: Teachers' Beliefs, Students' Engagement and Social Interaction (Research in Mathematics Education). Cham: Springer International Publishing AG, P. 253 – 263.

Girnat, 2018

Girnat, B. (2018). The PISA Mathematics Self-Efficacy Scale: Questions of Dimensionality and a Latent Class Concerning Algebra. In H. Palmér G J. Skott (Eds), Students' and Teachers' Values, Attitudes, Feelings and Beliefs in Mathematics Classrooms. Springer International Publishing AG.

Grob et al., 1991

Grob, A., Lüthi, R., Kaiser, F. G., Flammer, A., Mackinnon, A., & Wearing, A. J. (1991). Berner Fragebogen zum Wohlbefinden Jugendlicher (BFW). Diagnostica, 37(1), 66-75.

GSES

Jerusalem, M. & Schwarzer, R. (1999). Allgemeine Selbstwirksamkeitserwartung. In: R. Schwarzer & M. Jerusalem (Hg.). Skalen zur Erfassung von Lehrer und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschamlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin.

Schwarzer, R. (1999). General Perceived Self-Efficacy in 14 Cultures. In. Berlin: Freie Universität Berlin.

Schwarzer, R. (2014). Documentation of the General Self-Efficacy Scale. Retrieved from http://www.ralfschwarzer.de/.

Hascher, 2004

Hascher, T. (2004). Wohlbefinden in der Schule. Münster: Waxmann.

Hagenauer & Hascher, 2012 (modified)

Hagenauer, G., & Hascher, T. (2012). Erfassung kognitiver Regulationsstrategien bei Schulunlust. Empirische Pädagogik, 26(4), 452-478.

Hascher et al., 2019

Hascher, T., Brühwiler, C., & Girnat, B. (2019). Erläuterungen zu den Skalen des Kon-textfragebogens der ÜGK 2016 Mathematikteil: Theoretischer Hintergrund und Forschungsinteressen. http://uegk-schweiz.ch/uegk-2016-neu/

Hobza, et al., 2017

Hobza, V., Hamrik, Z., Bucksch, J. & De Clercq, B. (2017). 'The Family Affluence Scale as an Indicator for Socioeconomic Status: Validation on Regional Income Differences in the Czech Republic'. International journal of environmental research and public health 14 (12): 1540-1549.

see also Hartley, J. E. K., Levin, K., & Currie, C. (2016). A new version of the HBSC Family Affluence Scale - FAS III: Scottish Qualitative Findings from the International FAS Development Study. Child Indicators Research, 9: 233–245.

ICILS 2013

Jung, M., & Carstens, R. (2015). International Computer and Information Literacy Study: ICILS 2013 User Guide for the International Database. Amsterdam: International Association for the

see also Fraillon, J., Ainley, J., Schulz, W., Friedman, T., Gebhardt, E. (2014). Preparing for Life in a Digital Age. The IEA International Information Literacy Study International Report. Cham: Springer.

IGLU 2001

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (2007): Internationale Grundschul-Lese-Untersuchung 2001 (IGLU/PIRLS 2001). Version: 1. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_IGLU_2001_v1

ISSP 2003

ISSP Research Group. (2003). International Social Survey Programme: Social Relations and Support Systems - ISSP 2001. In. Köln: GESIS Datenarchiv.

PISA 2012

PISA Programme for International Student Assessment (2014). PISA 2012 technical report. Paris: OECD.

see also Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K., & González Rodŕıguez, E. (Eds.) (2018). PISA 2012 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster/New York: Waxmann.

PISA 2006

PISA Programme for International Student Assessment (2009). PISA 2006 technical report. Paris: OECD.

see also Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., . . . Pekrun, R. (Eds.) (2009). PISA 2006 Skalenhandbuch: Dokumentation der Erhebungsinstrumente. Münster: Waxmann.

PISA 2000

Adams, R. and Wu, M. (2002). PISA 2000 technical report. Paris: OECD.

see also Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., . . . Weiß, M. (Eds.) (2002). PISA 2000: Dokumentation der Erhebungsinstrumente. Berlin: Max-Planck-Institut für Bildungsforschung.

Rammstedt et al., 2014

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A., (2014). Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen. Doi:10.6102/zis76

see also Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41(1), 203-212. doi:10.1016/j.jrp.2006.02.001

see also NEPS. (2013). Startkohorte 4: Klasse 9 (SC4) Wellen 1 und 2 Erhebungsinstrumente (SUF-Version 1.1.0). In. Bamberg: Universität Bamberg, Nationales Bildungspanel (NEPS).

List of Sources (continued)

Rosenberg 1979 (translated)

Rosenberg, M. (1979). Conceiving the self. New York: Basic Books.

Ryan & Conell, 1989

Ryan, R., & Connell, J. (1989). Perceived locus of causality and internalization: Examing reasons for acting in two domains. Journal of Personality and Social Psychology, 57, 749-761.

SELLMO 2012

Spinath, B., Stiensmeier-Pelster, J., Schöne, C., & Dickhäuser, O. (2002). Skalen zur Erfassung der Lern- und Leistungsmotivation (SELLMO). Göttingen: Hogrefe.

Seidel, Prenzel & Kobarg, 2005

Seidel, T., Prenzel, M., & Kobarg, M. (Eds.). (2005). How to run a video study: Technical report of the IPN Video Study. Münster: Waxmann.

Sen, 1985; Anand & van Hees, 2006

Sen, A. K. (1985). Commodities and Capabilities, Oxford: Elsevier

Anand, P., & van Hees, M. (2006). Capabilities and achievements: An empirical study. The Journal of Socio-Economics, 35(2), 268-284.

Szydlik, 2008

Szydlik, M. (2008). Intergenerational Solidarity and Conflict. Journal of Comparative Family Studies, 39(1), 97-114.

TREE1

TREE (2016). Concepts and Scales. Survey waves 1 to 9, 2001-2014. Berne: TREE.

TREE2

Hupka-Brunner, S., Jann, B., Meyer, T., Imdorf, C., Sacchi, S., Müller, B., Scharenberg, K., von Rotz, C., Koomen, M. & Becker, R. (2015). Erläuterungen zum Kontextfragebogen der ÜGK 2016: Allgemeiner Teil. Bern.

Watermann, 2000

Watermann, R. (2000). Berufliche Wertorientierungen im Wandel. Eine Kohortenanalyse zur Dynamik arbeitsbezogener Einstellungen anhand von ALLBUS-Umfragedaten. Münster: Institut für sozialwissenschaftliche Forschung e.V

List of scales (wave 0)