Development and validation of the HLS-EU-Q12

Karin Waldherr, Tobias Alfers, Sandra Peer

Content

Vienna, September 2021

1. Background

The European Health Literacy Questionnaire (HLS-EU-Q47) consists of 47 items, which according to the underlying conceptual model, address a matrix of 3 by 4 domains resulting in 12 elements of the health literacy (HL) conceptual matrix (cf. Sørensen et al., 2013; Sørensen et al., 2015; Pelikan & Ganahl, 2017). Accordingly, the 47 items assess self-reported difficulties in the four cognitive domains accessing, understanding, appraising and applying information relevant for taking decisions in the three health domains healthcare, disease prevention and health promotion (Sørensen et al., 2013; Sørensen et al., 2015). Participants are asked to rate each item on a 4-point Likert like scale (very easy, fairly easy, fairly difficult, very difficult). Furthermore, they have the option to choose “don’t know”.

The items were developed in English and then translated into Bulgarian, Dutch, German, Greek, Polish and Spanish. The psychometric properties of the questionnaire was investigated using Principal Component Analysis (PCA) and reliability analysis using data from a field test conducted in Ireland and the Netherlands (for more details on the development process see Sørensen et al., 2013; Sørensen et al., 2015). The HLS-EU-Q47 was applied in the first wave of the European Health Literacy Survey in eight countries (HLS-EU-8): Austria (AT), Germany (only North-Rhine-Westphalia, DE), Spain (ES), Ireland (IE), The Netherlands (NL), Bulgaria (BG), Poland (PL), and Greece (EL). Data was collected either by Computer Assisted Personal Interviewing (CAPI) or Paper Assisted Personal Interviewing (PAPI). Recruitment strategies varied between countries (cf. Pelikan & Ganahl, 2017).

Using data from HLS-EU-8, four main index scores were constructed for “general HL” (comprising all 47 items), “healthcare literacy”, “disease prevention literacy” and “health promotion literacy”, and reliability for these indexes was assessed using Cronbach’s α. The Cronbach α’s for all four indexes across all eight countries were at least 0.87 and the item correlations with the total scales exceeded 0.30 (HLS-EU Consortium, 2012). Furthermore, in order to justify the usage of an overall sum score, Item Response Theory (IRT) analysis was applied to examine unidimensionality of the HLS-EU-Q47. The Rating Scale Model (RSM; Andrich, 1978) was used with the four-point scale and the Rasch Model (RM) with dichotomized data (very easy / fairly easy vs. fairly difficult / very difficult). The RSM analysis showed poor model fit. To test the fit of the RM to the data Likelihood Ratio Tests (Andersen, 1973) using the split criteria median test score, gender and dichotomized educational level were conducted for each of the eight countries. As result from these analyses a 16-item version was proposed (HLS-EU-Q16; only unpublished manuscript available; for more details see Pelikan & Ganahl, 2017). Correlations between the indexes of this short version and the 47-item version varied between r = 0.73 and r = 0.88 in the different countries (cf. Pelikan & Ganahl, 2017). However, the HLS-EU-Q16 does not include an item of the element „apply information” in the “health promotion“ domain of the HL conceptual matrix.

In the last years, both in Norway (HLS-Q12; Finbråten et al., 2017; Finbråten et al., 2018) and in Taiwan (HL-SF12; Duong et al., 2017) 12-item versions of the HLS-EU were developed in which each of the elements of the HL conceptual matrix is represented by one item. Whereas Finbråten et al. (2017) and Finbråten et al. (2018) applied Confirmatory Factor Analysis (CFA) and IRT, Duong et al. (2017) used only CFA to examine the psychometric properties of their 12-item version. However, only 50% of the items of these two 12-item versions are overlapping (see Table 1). Four of this six items are also contained in the HLS-EU-Q16.

Table 1: Items included in the HLS-EU-Q16, the HLS-Q12 (Finbråten et al., 2017; Finbråten et al., 2018) and the HL-SF12 (Duong et al., 2017)

HL conceptual matrix element	Item number in the HLS-EU-Q47		
	HLS-EU-Q16	HLS-Q12	HL-SF12
1 (acess information, healthcare)	2, 4	2	2
2 (understand information, healthcare)	5, 8	7	6
3 (appraise information, healthcare)	11	10	10
4 (apply information, healthcare)	13, 16	14	15
5 (acess information, disease prevention)	18	18	18
6 (understand information, disease prevention)	21, 23	23	23
7 (appraise information, disease prevention)	28	28	26
8 (apply information, disease prevention)	31	30	30
9 (acess information, health promotion)	33	32	33
10 (understand information, health promotion)	37, 39	38	39
11 (appraise information, health promotion)	43	43	43
12 (apply information, health promotion)	-	44	45

A short version representing all 12 elements of the HL conceptual matrix by one item which sufficiently meets the requirements of a unidimensional IRT model is highly desirable for several reasons. Therefore, in preparation of the second wave of the European Health Literacy Survey (HLS_{19}) new IRT-analyses using data from HLS-EU-8 were conducted with the goal to select a subsample of items - the HLS-EU-Q12 - which should fulfill the following criteria.

The HLS-EU-Q12 should

1. represent all 12 elements of the HL conceptual matrix by one item,
2. include as many items from the HLS-EU-Q16 as possible (cf. Table 1),
3. show the greatest possible overlap with the HLS-Q12 (Finbråten et al., 2017; Finbråten et al., 2018; cf. Table 1), and
4. represent a close to optimal 12-item solution, i.e. the solution with the lowest deviance from the assumptions of the Partial Credit Model (PCM; Masters, 1982) when analyzed separately for each HLS-EU-8 country.

In the following, the development of the HLS-EU-Q12 based on HLS-EU-8 data and its validation using data from HLS_{19} is described.

2. Development of the HLS-EU-Q12

2.1 Methods

Participants

Analyses are based on data from all eight countries of the HLS-EU-8 study collected in 2011. A detailed description of the HLS-EU-8 recruitment strategies in the different countries can be found elsewhere (e.g. Sørensen et al., 2015; Pelikan & Ganahl, 2017). Across all HLS-EU-8 countries data from n = 8102 persons were available whereby sample sizes varied between n = 1000 and n = 1057 in the individual countries (see Table 2).

Table 2: Sample sizes in the 8 HLS-EU countries

		Sample size
Country	AT	1015
	BG	1002
	EL	1000
	ES	1000
	IE	1005
	NL	1023
	PL	1000
	DE	1057
	Total	8102

Data analysis

The data set was divided randomly into a training data set (n = 4054) and a test data set (n = 4048). An iterative IRT analysis approach combined with expert judgement on content validity was chosen, including the following steps:

1) PCM analysis of HLS-EU-Q47 on the training data set across all HLS-EU-8 countries (n = 4054): The goal was to find additional items on top of HLS-EU-Q16 which could be used for item selection for the HLS-EU-Q12.
2) Selection of additional items based on the results of the PCM analysis (Step 1) and expert judgement on content validity (exclude low priority items).
3) PCM analysis of HLS-EU-Q16 plus additional items chosen in Step 2 using the test data set (n = 4048) for each HLS-EU-8 country separately: The aims of this step were to evaluate the item selection of Step 2 for each of the HLS-EU-8 country and to find the HLS-EU-Q12 solution with the best fit to the PCM.
4) PCM analysis of the selected 12 items (from Step 3) on the same test data set (n = 4048) for each HLS-EU-8 country. Since in Step 3 some items have been removed, the remaining 12 items were retested to evaluate if the scale has been affected (cf. Robinson et al., 2019).
5) Comparison of PCM model fit of the different questionnaire versions (HLS-EU-Q47, HLS-EU-Q16, HLS-EU-Q12, HLS-Q12, HL-SF12) using the test data set (n = 4048) for each HLS-EU-8 country, examination of the correlations of the HLS-EU-Q12 with the HLS-EU-Q47, the HLS-EU-

Q16 and the Newest Vital Sign test (NVS; Weiss et al., 2005[1]), and calculation of Cronbach's α as well as item-total correlations for the HLS-EU-Q12. In order to calculate the correlations of the HLS-EU-Q12 with the HLS-EU-Q47, HLS-EU-Q16 and the NVS, indices of HL were constructed as described in Sørensen et al. (2015).

PCM analysis:

All analyses were conducted in R 3.5.1 (https://cran.r-project.org/) using the packages TAM 3.1-45 (Robitzsch, Kiefer & Wu, 2019, 2020), sirt 3.3.-26 (Robitzsch, 2019), and mirt (Chalmers, 2012; version 1.30). Persons with more than 3 missing values were excluded. The PCM with ConQuest parametrization was used (Robitzsch, Kiefer & Wu, 2019).

In Steps 1, 3 and 4, item infit statistics and corresponding *t*-statistics were calculated for each item. The expected value is 1; values > 1 indicate that the item is less predictable than what would be expected according to the IRT model (underfit), values < 1 mean that the item is more predictable than what would be expected according to the expectations of the IRT model (= overfit; Linacre & Wright, 1994, p. 360). Underfitting items may severely degrade the measurement, whereas overfitting items may overestimate raw score differences (Smith et al., 2008). The Holm procedure was applied to adjust the *p*-values for multiple testing (cf. Robitzsch, Kiefer & Wu, 2019). Items were interpreted as over-/underfitting if the adjusted *p*-value was ≤ 0.05. The Nominal Categories Model was applied to check whether the expected ordering of response categories is supported by the data (Thissen, Cai & Bock, 2010; Chalmers et al., 2019, p. 100). Differential item functioning (DIF) analyses were conducted using gender and the dichotomized criteria age (median split) and education (< higher education entrance qualification vs. at least higher education entrance qualification). A facets analysis was conducted. The criteria were set up as facets (e.g. for gender, item+gender+item*gender), and the IRT analysis was rerun (Robitzsch, Kiefer & Wu, 2019). The interaction term item*gender yields the DIF magnitude.

For the comparison of PCM model fit of the different questionnaire versions (Step 5), *SRMSR* (standardized root mean square residual; Maydeu-Olivares, 2013) was calculated for each of the questionnaire versions (cf. Robitzsch, Kiefer & Wu, 2019). *SRMSR* is a global fit statistic based on the comparison of residual correlations of item pairs. Maydeu-Olivares suggests a cutoff of ≤ 0.05 for well-fitting IRT models. A less conservative value of 0.08 often is used as acceptable (cf. Hu & Bentler, 1999). Furthermore, the combined PCA / t-test protocol to examine unidimensionality (cf. Smith, 2002; Hagell, 2014) was applied to the different versions. Two subsets of items are formed based on a PCA of standardized item residuals pursuant to the loadings of the item residuals on the first principal component (cf. Hagell, 2014). Person parameter estimation is conducted in each of the two item subsets and the resulting person parameter estimates from the two subsets are compared by means of paired *t*-tests (cf. Hagell, 2014). Under the assumption of unidimensionality, the proportion of individuals with significantly different person parameters in the two item subsets is small, i.e. ≤ 5% of the *t*-tests are significant, or the lower bound of a 95% confidence interval (CI) of the observed proportion overlaps 5% (Hagell, 2014). In our analysis the Agresti-Coull CI was used. *WLE* reliability and *EAP* (expected a posteriory) reliability coefficients were calculated according to Adams (2005) (cf. Robitzsch, Kiefer & Wu, 2019). Additionally, deviance, Akaike's Information Criterion (*AIC*; Akaike, 1973), the *AIC* correction for small samples (*AICc*; Hurvich & Tsai, 1989), Bozdogan's (1987) consistent

[1] The NVS is a 6-item screening instrument for functional health literacy and is based on the ability to read, understand and apply information from a nutrition label.

AIC (*CAIC*) and the Bayesian Information Criterion (*BIC*; Schwarz, 1978) were calculated to compare the data-model fit for the different versions. Lower values indicate better data-model fit.

2.2 Results

Step 1:

No unordered response categories were observed. Seven items of the HLS-EU-Q16 had significant infit statistics (see Table A1 in the Appendix). Overfit was observed for six items (items 13, 21, 23, 33, 39, 43), and underfit was observed for item 28 with an infit statistic of 1.10 (t = 4.60, p < 0.001). Another 17 items of the remaining items of the HSL-EU-Q47 had significant infit statistics. DIF was observed for 12 items of the HLS-EU-Q16 (items 5, 18, 39 for age; items 8, 21, 23, 28, 31, 37 for education, and items 2, 11 and 33 for age and education). For 6 items of the HLS-EU-Q47 which are not included in the HLS-EU-Q16 neither over-/underfit nor DIF was observed (see Table A1). As in previous analyses, the most problematic subdomain was “health promotion”; only for two items of this subdomain neither DIF nor over-/underfit was observed.

Step 2:

Six items were candidates to be selected as additional items on top of the 16 items of the HLS-EU-Q16 according to the results of Step 1. Two of them were judged as low priority items and were not considered. Thus, four items were selected: 7, 10, 24, 44. Furthermore, it was decided to include two additional items from the health promotion domain, although they showed DIF in the training data set (items 36 and 42), such that each of the four cognitive domains (access, understand, appraise and apply) was represented by two items. This resulted in six additional items, whereby three of them are included in the HLS-Q12 (Finbråten et al., 2017).

Step 3:

Using the test data set, only for items 28 (infit: 1.23, t = 3.47, p = 0.032) and 36 (infit: 1.46, t = 5.75, p < 0.001) significant underfit was observed in Germany (see Table A2 in the Appendix). DIF for age was observed for item 2 in four countries (AT, EL, ES, NL), for item 33 in two countries (BG, EL), and for item 23, 39 and 42 in one country. DIF for gender was found for items 5 and 43 in one country, and DIF for education was found for item 8 in two countries as well as for items 11, 21 and 31 in one country.

The proposed solution fulfilling the abovementioned criteria (represent all 12 elements of the HL conceptual matrix by one item, include as many items of the HLS-EU-Q16 as possible, show the greatest possible overlap with the HLS-Q12 and showing the lowest deviance from the assumptions of the PCM across all HLS-EU-8 countries) consisted of the items 4, 7, 10, 16, 18, 23, 24, 31, 33, 37, 42, 44.

Step 4:

PCM analysis of the selected 12 items in each of the 8 countries revealed no significant infit statistics, however DIF for education for item 31 in Austria and for item 33 in Bulgaria, and DIF for age for item 33 in Bulgaria and Greece as well as for item 42 in The Netherlands (see Table A3 in the Appendix).

Step 5:

SRMSR-values for the HLS-EU-Q12 were < 0.08 in the individual countries and thus are acceptable (Table 3). For the HLS-Q12 values > 0.08 were observed in two countries and for the HL-SF12 in three countries.

Table 3: SRMSR values for the different versions in the eight countries

Country	HLS-EU-Q47	HLS-EU-Q16	HLS-EU-Q12	HLS-Q12	HL-SF12
AT	**0.0920**	**0.0828**	0.0712	0.0768	0.0775
BG	**0.0891**	0.0773	0.0696	0.0713	**0.0900**
EL	**0.1031**	**0.0912**	0.0769	**0.0936**	**0.1031**
ES	**0.0920**	0.0748	0.0683	0.0653	0.0760
IE	**0.0949**	**0.0885**	0.0789	0.0776	0.0716
NL	**0.0947**	**0.0885**	0.0745	0.0742	0.0767
PL	0.0795	0.0643	0.0595	0.0543	0.0644
DE	**0.1129**	**0.0921**	0.0798	**0.1105**	**0.0898**

For each of the three 12-item versions the proportions of significant *t*-tests were > 5% in all countries, and only in one country the lower bound of the 95% CI included 5% for each of the versions (Table 4). In three countries the proportion of significant *t*-tests was lowest for the HLS-EU-Q12 (ES, IE, NL), and in one country for the HL-SF12 (BG). In AT, EL and DE the proportion was comparable for the HLS-EU-Q12 and HL-SF12, and in PL it was comparable for the HLS-Q12 and HL-SF12.

Table 4: Results of PCA/t-test procedure (proportion of significant t-tests and lower bound of CI)

Country	HLS-EU-Q47	HLS-EU-Q16	HLS-EU-Q12	HLS-Q12	HL-SF12
AT	**0.259 (0.224)**	**0.184 (0.154)**	**0.086 (0.065)**	**0.121 (0.096)**	**0.082 (0.061)**
BG	**0.310 (0.271)**	**0.158 (0.129)**	**0.125 (0.099)**	**0.104 (0.080)**	**0.087 (0.066)**
EL	**0.271 (0.234)**	**0.159 (0.130)**	**0.101 (0.077)**	**0.114 (0.089)**	**0.098 (0.076)**
ES	**0.253 (0.217)**	**0.179 (0.147)**	**0.101 (0.077)**	**0.113 (0.088)**	**0.143 (0.115)**
IE	**0.289 (0.251)**	**0.139 (0.111)**	**0.084 (0.062)**	**0.102 (0.078)**	**0.104 (0.080)**
NL	**0.216 (0.182)**	**0.141 (0.112)**	0.069 (0.049)	**0.107 (0.082)**	**0.074 (0.054)**
PL	**0.234 (0.199)**	**0.103 (0.079)**	**0.078 (0.057)**	0.066 (0.047)	0.063 (0.044)
DE	**0.291 (0.253)**	**0.146 (0.118)**	**0.078 (0.058)**	**0.143 (0.115)**	**0.076 (0.056)**

When comparing the three 12-item versions by means of deviance and information criteria, the HLS-EU-Q12 showed best fit to the PCM (i.e. consistently had the lowest values in seven of the eight countries); in Austria the HL-SF12 had the lowest values (see Table A4 in the Appendix). All three 12-item versions had acceptable *WLE* and *EAP* reliability coefficients > 0.77 in all eight countries (see Table 5).

Table 5: WLE and EAP reliability coefficients for the different questionnaire versions

Country	HLS-EU-Q47	HLS-EU-Q16	HLS-EU-Q12	HLS-Q12	HL-SF12
WLE reliability coeff.					
AT	0.948	0.868	0.830	0.828	0.835
BG	0.965	0.912	0.884	0.884	0.882
EL	0.952	0.886	0.846	0.851	0.850
ES	0.951	0.873	0.830	0.837	0.827
IE	0.945	0.874	0.838	0.839	0.839
NL	0.938	0.839	0.781	0.784	0.771
PL	0.962	0.901	0.870	0.881	0.873
DE	0.946	0.878	0.829	0.827	0.820
EAP reliability coeff.					
AT	0.955	0.877	0.840	0.836	0.845
BG	0.972	0.927	0.902	0.897	0.897
EL	0.965	0.906	0.867	0.872	0.871
ES	0.954	0.876	0.835	0.841	0.829
IE	0.960	0.897	0.868	0.864	0.865
NL	0.944	0.861	0.813	0.807	0.800
PL	0.970	0.917	0.893	0.898	0.889
DE	0.958	0.905	0.860	0.850	0.850

The correlation of the HLS-EU-Q12 and the HLS-EU-Q47 indices was high in the total sample of all eight countries (r = 0.957). In the individual countries the correlations varied between 0.938 and 0.967 (see Table 6). The correlations with the HLS-EU-Q16 were comparable. The correlation of the HLS-EU-Q12 index with the NVS was r = 0.26 in the total sample and the correlations in the individual countries varied between r = 0.13 and r = 0.269. These values are comparable to the correlations of the HLS-EU-Q47 index with the NVS (r = 0.25 for the total EU-8, and correlations between r = 0.14 and r = 0.38 in the individual countries; cf. Pelikan & Ganahl, 2017).

Table 6: Correlations of the HLS-EU-Q12 with HLS-EU-Q16 and NVS

	AT	BG	EL	ES	IE	NL	PL	DE	Total (EU-8)
HLS-EU-Q47	0,946	0,967	0,960	0,938	0,963	0,938	0,962	0,961	0,957
HLS-EU-Q16	0,930	0,970	0,952	0,931	0,953	0,929	0,965	0,945	0,951
NVS	0,153	0,399	0,322	0,210	0,269	0,190	0,392	0,130	0,26

Replacing item 33 by item 32 and evaluating model fit

Following a consortium decision, it was examined if item 33 could be replaced by item 32. Therefore, Steps 4 and 5 were applied to a version consisting of items 4, 7, 10, 16, 18, 23, 24, 31, 32, 37, 42, 44 (called HLS-EU-Q12$_{32}$ in the following) in order to evaluate its model fit.

For the HLS-EU-Q12$_{32}$ no significant infit statistics were observed (see Table A5 in the Appendix), as was the case for the version containing item 33 instead of item 32. DIF for age was observed for item 32 in two countries (BG, EL) and DIF for education in two countries (BG, IE). In three countries *SRSMR*-values > 0.08 were observed for the HLS-EU-Q12$_{32}$, and the proportion of significant t-tests was < 5% only in one country (see Table 7). *WLE* und *EAP* reliability coefficients were comparable for both test versions with values > 0.77 for HLS-EU-Q12$_{32}$ and > 0.78 for the HLS-EU-Q12 in all countries. Comparing the two versions by deviance and information statistics, the version containing item 32 shows consistently lower values across all statistics and across all countries.

Table 7: Comparison of HLS-EU-Q12 und HLS-EU-Q12$_{32}$

	Country							
	AT	**BG**	**EL**	**ES**	**IE**	**NL**	**PL**	**DE**
SRMSR								
HLS-EU-Q12	0.0712	0.0696	0.0769	0.0683	0.0789	0.0745	0.0595	0.0798
HLS-EU-Q12$_{32}$	0.0728	0.0699	**0.0806**	0.0693	**0.0852**	0.0739	0.0631	0.0881
PCA/t-test (proportion significant t-tests, CI)								
HLS-EU-Q12	**0.086 (0.065)**	**0.125 (0.099)**	**0.101 (0.077)**	**0.101 (0.007)**	**0.084 (0.062)**	0.069 (0.049)	**0.078 (0.057)**	**0.078 (0.058)**
HLS-EU-Q12$_{32}$	**0.095 (0.073)**	**0.079 (0.058)**	**0.122 (0.096)**	**0.079 (0.058)**	**0.104 (0.08)**	0.039 (0.025)	**0.086 (0.064)**	**0.104 (0.08)**
***WLE* reliability coeff**								
HLS-EU-Q12	0.8304	0.8838	0.8464	0.8298	0.8383	0.7806	0.8703	0.8293
HLS-EU-Q12$_{32}$	0.8271	0.8842	0.8444	0.8315	0.8357	0.7761	0.8710	0.8308
***EAP* reliability coff.**								
HLS-EU-Q12	0.8304	0.8838	0.8464	0.8298	0.8383	0.8137	0.8933	0.8601
HLS-EU-Q12$_{32}$	0.8368	0.9010	0.8654	0.8365	0.8666	0.8112	0.8939	0.8622

Deviance								
HLS-EU-Q12	13130.16	11309.64	11713.98	9370.28	10756.19	10749.74	9835.49	12001.58
HLS-EU-Q12$_{32}$	13074.86	11294.01	11694.90	9269.33	10570.27	10516.50	9792.51	11895.35
AIC								
HLS-EU-Q12	13204.16	11383.28	11787.98	9444.56	10830.19	10823.74	9909.49	12075.58
HLS-EU-Q12$_{32}$	13148.86	11368.01	11768.90	9343.33	10644.27	10590.50	9866.51	11969.35
AICc								
HLS-EU-Q12	13209.79	11389.25	11793.86	9450.79	10836.39	10829.85	9915.67	12081.51
HLS-EU-Q12$_{32}$	13154.49	11373.98	11774.77	9349.55	10650.47	10596.61	9872.69	11975.28
CAIC								
HLS-EU-Q12	13399.74	11576.88	11982.16	9636.75	11022.54	11016.53	10101.91	12269.40
HLS-EU-Q12$_{32}$	13344.44	11561.61	11963.08	9535.52	10836.62	10783.29	10058.93	12163.17
BIC								
HLS-EU-Q12	13362.74	11539.88	11945.16	9599.75	10985.54	10979.53	10064.91	12232.40
HLS-EU-Q12$_{32}$	13307.44	11524.61	11926.08	9498.52	10799.62	10746.29	10021.93	12126.17

The correlation of the HLS-EU-Q12$_{32}$ and the HLS-EU-Q47 indices was r = 0.955 in the total sample of all eight countries and therefore comparable with the correlation of the version containing item 33. In the individual countries the correlations varied between r = 0.935 and r = 0.966. The correlations with the HLS-EU-Q16 were also comparable. The correlation of the HLS-EU-Q12 index with the NVS was r = 0.263 in the total sample and the correlations in the individual countries varied between r = 0.13 and r = 0.385. Therefore, all correlations are comparable with the version containing item 33 instead of item 32.

Table 8: Correlations of the HLS-EU-Q12$_{32}$ with HLS-EU-Q16 and NVS

	AT	BG	EL	ES	IE	NL	PL	DE	Total (EU-8)
HLS-EU-Q47	0.949	0.966	0.959	0.938	0.954	0.935	0.956	0.959	0.955
HLS-EU-Q16	0.929	0.969	0.945	0.925	0.942	0.919	0.959	0.940	0.946
NVS	0.168	0.406	0.314	0.211	0.266	0.195	0.385	0.130	0.263

3. The HLS$_{19}$-Q12

According to a consortium decision, in HLS$_{19}$ the 12-item version containing items 4, 7, 10, 16, 18, 23, 24, 31, 32, 37, 42, 44, with some improvement in the wording of a few items, and a change in the wording of the response categories (omitting the qualifier “fairly” in the two middle categories, resulting in the four point scale “very easy”, “easy”, “difficult”, “very difficult”) were used and named HLS$_{19}$-Q12. Figure 1 shows the instruction and items of the HLS$_{19}$-Q12.

Figure 1: Instruction and items of the HLS$_{19}$-Q12

INTRODUCTION	***Interviewer: It is not always easy to get understandable, reliable and useful information on health related to*** ***With the following questions we would like to find out which tasks related to handling health information ar difficult.***
CORE-HL	On a scale from very easy to very difficult, how easy would you say it is:
	[SHOWCARD WITH SCALE - ONE ANSWER PER ROW]

Very easy	Easy	Difficult

CORE-HL4	...to find out where to get professional help when you are ill? ***[instructions: such as doctor, nurse, pharmacist, psychologist]***
CORE-HL7	...to understand information about what to do in a medical emergency?
CORE-HL10	...to judge the advantages and disadvantages of different treatment options?
CORE-HL16	...to act on advice from your doctor or pharmacist?
CORE-HL18	...to find information on how to handle mental health problems? ***[Instruction: stress, depression or anxiety]***
CORE-HL23	...to understand information about recommended health screenings or examinations? ***[Instructions: e.g. colorectal cancer screening, blood sugar test]***
CORE-HL24	...to judge if information on unhealthy habits, such as smoking, low physical activity or drinking too much alcohol, are reliable?
CORE-HL31	...to decide how you can protect yourself from illness using information from the mass media? ***[Instructions: e.g. Newspapers, TV or Internet]***
CORE-HL32	...to find information on healthy life styles such as physical exercise, healthy food or nutrition?
CORE-HL37	...to understand advice concerning your health from family or friends?
CORE-HL42	...to judge how your housing conditions may affect your health and well-being?
CORE-HL44	...to make decisions to improve your health and well-being?

4. Validation of the HLS_{19}-Q12

The aim was to evaluate the PCM model fit of the HLS_{19}-Q12 using the data of the second HLS-EU wave. Furthermore, it was aimed to test the model fit of the dichotomized version to the Rasch model.

3.1 Methods

Participants

Analyses are based on data from 15 countries of the second wave of the HLS-EU study collected between November 2019 and February 2021. A detailed description of the recruitment strategies in the different countries can be found in Chapter 2 of the HLS_{19} International Report.

Table 9 gives an overview on the data collection methods, sample sizes and version of the questionnaire used (all 47 items of the HLS-EU-Q47, the HLS-Q16 plus 6 items on top selected in step 2 of the HLS_{19}-Q12 development, or only the 12 items of the HLS_{19}-Q12) for the individual countries. Across all countries data from n = 38080 persons were available whereby sample sizes varied between n = 1000 and n = 5660 in the individual countries.

Table 9: Overview on analyzed data for the validation of the HLS_{19}-Q12

Country	Data collection method	HLS version	N	N without missing items
Austria (AT)	CATI	HLS_{19}-Q12	2967	2471
Belgium (BE)	CAWI	HLS-Q16 plus 6	1000	1000
Czech Republic (CZ)	CATI, CAWI	HLS-Q16 plus 6	1599	1459
Denmark (DK)	CAWI	HLS-Q16 plus 6	3602	3506
Germany (DE)	PAPI	HLS-EU-Q47	2143	1991
France (FR)	CAWI	HLS-Q16 plus 6	2003	2003
Hungary	CATI	HLS-Q16 plus 6	1195	1021
Ireland (IE)	CATI	HLS-EU-Q47	4487	4142
Israel (IL)	CATI, CAWI	HLS-Q16 plus 6	1315	1294
Norway (NO)	CATI	HLS-EU-Q47	2855	2387
Portugal (PT)	CAWI	HLS_{19}-Q12	1247	922
Russia (RU)	PAPI	HLS-Q16 plus 6	5660	4752
Slovenia (SI)	CAPI, Paper, CAWI	HLS-EU-Q47	3360	3178
Slovakia (SK)	CAPI	HLS-Q16 plus 6	2145	2144
Switzerland (CH)	CATI, CAWI	HLS_{19}-Q12	2502	2370

Data analysis

The assumption of unidimensionality for HLS_{19}-Q12 was tested by means of the PCM in a first step and the dichotomous Rasch Model (RM) in a subsequent step. All analyses were conducted in R using the packages eRm 1.0-1 (Mair et al., 2018), TAM 3.5-19 (Robitzsch, Kiefer & Wu, 2020) and mirt 1.33.2 (Chalmers, 2015). Persons with missing values on at least one of the items were excluded. Analyses were conducted for each of the countries separately. Due to very large sample sizes in some countries (e.g. RU and IE), all analyses were conducted also in a random sample of n=900 for each of the countries (the sample size for the random sample n=900), and PCM analyses on item level were also conducted in four randomly chosen independent subsamples in each of the countries (therefore the sample sizes in the four subsamples vary according to the total sample sizes in the individual countries). Due to the huge number of significance tests and large sample sizes, $\alpha = 0.001$ was chosen.

For PCM analysis the same methods were applied as in the development of the HLS_{19}-Q12 described in 2.1:

- Individual items:
    - item infit statistics and corresponding *t*-statistics were calculated for each item;
    - the Nominal Categories Model was applied to check whether the expected ordering of response categories is supported by the data;
    - DIF analyses were conducted using the split criteria gender, median age and education (< higher education entrance qualification vs. at least higher education entrance qualification)
- Global fit statistics:
    - *SRMSR* (standardized root mean square residual;
    - combined PCA / t-test protocol;
    - *WLE* reliability and *EAP* (expected a posteriori) reliability coefficients.

Additionally, local stochastic independence was assessed by means of an adjusted variant of the *Q3* statistic by Yen (1984), *aQ3*, for all item pairs and an effect size of model fit (*MADaQ3*), which is the average of the absolute values of *aQ3* statistics and *p*-values adjusted according to the Holm procedure (Robitzsch, Kiefer & Wu, 2020). For those countries, in which different data collection methods were applied (CH, CZ, IL, SI), analyses were conducted both independent of survey-type as well as separately for the survey-types. However, sample sizes were very small for CATI especially in CH (n=139). Therefore, no additional analyses were performed with random samples in survey-type-specific analyses.

For the dichotomous scoring, Likelihood-Ratio-Tests (LR-test, Andersen, 1973) were conducted as global model tests using median test score, education (< higher education entrance qualification vs. at least higher education entrance qualification), median age and gender as split criteria, and individual item-fit statistics (Fischer-Scheiblechner *z*-statistics, Fischer & Scheiblechner, 1970) were calculated. A global test for local independence, which calculates the sum of absolute deviations between the observed inter-item correlations and the expected correlations (Mair et al., 2020), was conducted. Furthermore, on item level increased correlations between inter-item residuals were checked by means of the *Q3*-statistic (cf. Mair et al., 2020). RM analyses were conducted in the total sample and the random sample of n=900. Furthermore, graphical model tests according to Rasch (1960/1980) were applied to examine model fit for each individual item. These model tests rely on the assumption that item parameters can be consistently estimated in different subsamples drawn from a population in which the Rasch model applies. Additionally, item characteristic curve (ICC) plots were used to graphically inspect model fit of the individual items. The ICC plots show how the probability for response category 1 expected by the RM changes with the values of the latent variable. Observed scores are represented by circles. If the deviations of the observed values from the expected values are small, there is close conformity of the data with the model.

3.2 Results

PCM analysis

Global fit statistics:

The *WLE* and *EAP* reliability coefficients have acceptable values > 0.78 in all countries and all survey types (cf. Table 10). For those countries which were included in the development of the HLS_{19}-Q12, the coefficients are comparable in AT, however lower in IE and DE (cf. Table 7, HLS-EU-$Q12_{32}$). Furthermore, Cronbach α's are sufficient in all countries and all survey types with values > 0.80.

Table 10: WLE and EAP reliability coefficients of the HLS_{19}-Q12

Country	WLE rel.	EAP rel.	Cronbach α	N
AT	0.830	0.850	0.843	2471
BE	0.879	0.883	0.881	1000
CH_Total	0.839	0.844	0.836	2370
CH_CAWI	0.841	0.847	0.838	2231
CH_CATI	0.811	0.808	0.801	139
CZ_Total	0.845	0.850	0.842	1459
CZ_CAWI	0.848	0.852	0.844	1057
CZ_CATI	0.820	0.823	0.814	402
DE	0.811	0.809	0.802	1991
DK	0.853	0.862	0.857	3506
FR	0.878	0.894	0.887	2003
HU	0.841	0.845	0.843	1021
IE	0.785	0.822	0.822	4142
IL_Total	0.874	0.890	0.882	1294
IL_CAWI	0.869	0.879	0.871	1004
IL_CATI	0.855	0.889	0.887	290
NO	0.831	0.844	0.838	2387
PT	0.839	0.853	0.902	922
RU	0.884	0.892	0.900	4752

Country	WLE rel.	EAP rel.	Cronbach α	N
SI_Total	0.873	0.887	0.892	3178
SI_CAWI	0.847	0.864	0.864	1463
SI_CAPI	0.887	0.899	0.912	1704
SK	0.881	0.887	0.884	2144

According to the PCA/t-test procedure, the HLS_{19}-Q12 cannot be considered sufficiently unidimensional except for NO, as the proportion of individuals with significant different person parameters in two item subsets exceeds 5% in all countries except for Norway in the random sample, and only for Norway the lower bound of the confidence interval (CI) overlaps 5% also in the total sample (see Table 11). In IE the lower bound of the CI is < 0.06 both in the total and the random sample, in IL in the total sample and in RU in the random sample, and in FR and CZ the lower bound of the CI is < 0.07 in the total sample. The highest percentages are observable in HU, PT and SK both in the total and the random samples. For SI results for the two survey modes are comparable; regarding the results for CATI in CH and IL it has to be considered that the samples sizes are small (n=139 and n=290, respectively). For those countries which were included in the development of the HLS_{19}-Q12 (AT, IE, DE), the results are slightly better or comparable than in the previous analysis (cf. Table 7, HLS-EU-$Q12_{32}$).

Table 11: Results of SRMSR and PCA/t-test procedure for the HLS_{19}-Q12

Country	*SRMSR*	PCA/t-test	
		Total sample	Random sample
AT	0.0596	0.093 (0.082)	0.077 (0.061)
BE	0.0659	0.088 (0.072)	0.093 (0.076)
CH_Total	0.0641	0.089 (0.078)	0.111 (0.092)
CH_CAWI	0.0651	0.096 (0.084)	-
CH_CATI	0.1104	0.165 (0.112)	-
CZ_Total	0.0541	0.080 (0.067)	0.104 (0.086)
CZ_CAWI	0.0559	0.092 (0.076)	-
CZ_CATI	0.0777	0.117 (0.089)	-
DE	0.0679	0.091 (0.080)	0.110 (0.091)
DK	0.0566	0.112 (0.102)	0.108 (0.089)
FR	0.0591	0.074 (0.063)	0.086 (0.069)
HU	0.0781	0.146 (0.126)	0.141 (0.120)
IE	0.0699	0.061 (0.054)	0.069 (0.054)

IL_Total	0.0485	0.070 (0.057)	0.083 (0.070)
IL_CAWI	0.0538	0.088 (0.072)	-
IL_CATI	0.0789	0.072 (0.047)	-
NO	0.0632	0.057 (0.049)	0.021 (0.013)
PT	0.0753	0.140 (0.119)	0.124 (0.104)
RU	0.0504	0.090 (0.082)	0.068 (0.053)
SI_Total	0.0775	0.085 (0.075)	0.113 (0.094)
SI_CAWI	0.0784	0.090 (0.076)	-
SI_CAPI	0.0672	0.090 (0.078)	-
SK	0.0556	0.116 (0.103)	0.128 (0.107)

The *SRMSR* statistics are above the cut-off value of 0.05 for good model fit suggested by Maydeu-Olivares (2013) in all countries except IL (see Table 11). However, the values are below the less conservative cut-off value of 0.08 according to Hu & Bentler (1999). But the global test for local independence based on the adjusted *Q3*-statistic yielded significant results in all countries both in the total and the random samples (see Table 12). For CATI the *Q3*-statistic was not significant in CH and IL; however, the sample sizes are very small.

Table 12: Results of the model tests for local independence based on the adjusted Q3-statistic in all countries in the total and the random samples

	Total Sample			**Random Sample**		
Country	***MADaQ3***	***max. aQ3***	***p***	***MADaQ3***	***max. aQ3***	***p***
AT	0.061	0.255	<0.001	0.058	0.246	<0.001
BE	0.070	0.200	<0.001	0.073	0.200	<0.001
CH_Total	0.061	0.287	<0.001	0.073	0.274	<0.001
CH_CAWI	0.061	0.286	<0.001	-	-	-
CH_CATI	0.108	0.309	0.012	-	-	-
CZ_Total	0.045	0.183	<0.001	0.050	0.187	<0.001
CZ_CAWI	0.051	0.223	<0.001	-	-	-
CZ_CATI	0.060	0.209	0.001	-	-	-
DE	0.067	0.170	<0.001	0.069	0.201	<0.001
DK	0.060	0.160	<0.001	0.062	0.172	<0.001
FR	0.061	0.266	<0.001	0.059	0.300	<0.001
HU	0.071	0.252	<0.001	0.070	0.235	<0.001

	Total Sample			Random Sample		
Country	*MADaQ3*	*max. aQ3*	*p*	*MADaQ3*	*max. aQ3*	*p*
IE	0.061	0.249	<0.001	0.071	0.282	<0.001
IL_Total	0.060	0.200	<0.001	0.063	0.221	<0.001
IL_CAWI	0.066	0.200	<0.001	-	-	-
IL_CATI	0.071	0.201	0.036	-	-	-
NO	0.053	0.157	<0.001	0.056	0.183	<0.001
PT	0.078	0.299	<0.001	0.077	0.304	<0.001
RU	0.068	0.220	<0.001	0.066	0.244	<0.001
SI_Total	0.059	0.244	<0.001	0.067	0.306	<0.001
SI_CAWI	0.052	0.212	<0.001	-	-	-
SI_CAPI	0.066	0.258	<0.001	-	-	-
SK	0.064	0.221	<0.001	0.072	0.284	<0.001

Analyses at item level:

The *Q3*-statistics for the item pairs in the random samples showed that the residuals of 14 item pairs are significantly correlated in several countries (see Table 13, Figure 2: Matrix of dependent item pairs and domains), also if the different survey modes are considered separately. Non significant results for CATI in CH, CZ and IL are partly due to the small sample sizes.

Table 13: Dependent item pairs in the different countries according to the random samples

Item pair	**Countries**
4 (access, HC) – 7 (understand, HC)	AT, CH[g], CZ[h], DE, DK, NO, PT[a,d], SI[h], SK[c]
10 (appraise, HC) – 32 (access, HP)	AT[b,e], BE[e], CH[b,f,h], CZ[c,f,g], DE[f], DK[c,f], IL[c,f,g], SK[c,f]
23 (understand, DP) – 24 (appraise, DP)	DE, FR[a,f], HU[d], IE, IL[g], RU[b], SI[a,e,h]
7 (understand, HC) – 10 (appraise, HC)	BE, CH[g], FR, IL[g], RU[c], SK
24 (appraise, DP) – 32 (access, HP)	AT, CH[g], CZ[g], DK, PT[a,d], SI[h]
42 (appraise, HP) – 44 (apply, HP)	AT, CH[b,f,g], IE[c], RU, SK
16 (apply, HC) – 18 (access, DP)	BE[c,f], CZ[g], PT
16 (apply, HC) – 31 (apply, DP)	BE[c,f], DE, PT
32 (access, HP) – 37 (understand, HP)	IL[h], PT[a,d], SK
4 (access, HC) – 31 (apply, DP)	BE[b,e], DE[c]
4 (access, HC) – 42 (appraise, HP)	BE[c,f], DE[c,f]
16 (apply, HC) – 32 (access, HP)	BE, IL[g], PT[b,e]
37 (understand, HP) – 42 (appraise, HP)	IE, SK, PT[d]

HC: healthcare; DP: disease prevention; HP: health promotion; [a]: r > 0.30 in the random sample; [b]: r > 0.25 in the random sample; [c]: r > 0.20 in the random sample; [d]: r > 0.30 in the total sample; [e]: r > 0.25 in the total sample; [f]: r > 0.20 in the total sample; [g]: total sample and CAWI; [h]: in all survey modes;

Two dependent item pairs were observed in the healthcare (HC) domain, one in the disease prevention (DP) domain and three in the health promotion (HP) domain in several countries and with a correlation of the residuals of r > 0.20 in at least one country in the total and/or the random sample. In the cognitive domain “apply information” one dependent item pair was observed. The remaining six dependent item pairs appeared across the domains.

Some more dependent item pairs were found only in PT or SK with correlations r > 0.20, with some correlations in PT above 0.30.

Figure 2: Matrix of dependent item pairs and domains

		HC			**DP**				**HP**			
	Items	**7**	**10**	**16**	**18**	**23**	**24**	**31**	**32**	**37**	**42**	**44**
HC	**4**	AT, CH, CZ, DE, DK, NO, PT, SI, SK						BE, DE		PT	BE, DE	
	7		BE, CH, FR, IL, RU, SK								SK	SK
	10								AT, BE, CH, CZ, DE, DK, IL, SK			
	16				BE, CZ, PT			BE, DE, PT	BE, IL, PT	PT		
DP	**23**						DE, FR, HU, IE, IL, RU, SI	SI				
	24								AT ,CH, CZ, DK, PT, SI	PT		
	31											
HP	**32**									PT, SK	PT	
	37										IE, PT, SK	
	42											AT, CH, IE, RU, SK

Table 14 shows the results for the item infit statistics, for the DIF analyses (only significant results are included in the table), and for the Nominal Response Model to identify the empirical ordering of the response categories. For Item 31 values of the infit statistic between 1.15 and 1.35, with high values of the corresponding *t*-statistics and *p*-values < 0.001, were observed in the total samples in four countries (CH, IE, NO, SI). In SI (infit statistic = 1.36, *t*=6.99, *p*<0.001) Item 31 displays also significant underfit in the random sample of n=900 and in all four independent subsamples as well as for both survey types. In IE (infit statistic = 1.19, *t*=4.23, *p* = 0.001) and NO (infit statistic = 1.18, *t*=3.76, *p*=0.006) the infit statistics are also high in the random sample of n=900, and in CH for CAWI (infit statistic 1.15, *t*=5.07, p<0.001). Furthermore, Item 4 shows significant underfit in FR in the total sample and the random sample of n=900. Item 23 displays underfit in IE, Item 37 in DE in the total sample, and Item 44 in SI in the total sample and the CAPI sample. All other significant infit statistics indicate overfit of the respective items which is, however, less problematic (for the infit statistics of all items in the total samples of the different countries see Table A6 in the Appendix).

Eleven of the items display DIF in at least one country for one split criterion in the total samples, and several items display DIF in more than one country also in the random sample of n=900 and/or at least one of the four independent subsamples (see

Table 14):

- Item 4 for
    - gender in three countries (AT, DE, IE), and
    - age in one country (AT)
- Item 7 for
    - age in two countries (DK, IE), and
    - education in IE,
- Item 10 for
    - education in three countries (AT, DK, SI[2])
- Item 16 for
    - gender in IE
- Item 18 for
    - age in IE, and
    - education in RU
- Item 23 for
    - gender in three countries (DK, IE, SI[3]),
    - age in six countries (AT, BE, CH[4], DK, FR, SI[3]) and
    - education in two countries (DK, SI[2]),
- Item 24 for
    - age in IE, underfit SI
- Item 31 for
    - gender in one country (SI[4]),
    - education in five countries (AT, CZ[2], IE, NO, SI[4]), and
    - age in one two countries (BE, SI[2])
- Item 32 for
    - gender in one country (DK),
    - age in three countries (CZ[3], RU, SI[3]) and
    - education in five countries (AT, CZ[4], FR, SI[3], SK),
- Item 37 for
    - age in four countries (AT, CH[4], DK, RU),
    - education in DE, and
- Item 42 for
    - age in four countries (CH[2], DK, IE, NO), and
    - education in IE.

[2] In the analysis including both survey types
[3] In all survey modes
[4] In the analysis including both survey types and for CAWI

Across countries, the most problematic items are items 4, 23, 31, 32, 37 and 42 both in the total and the random samples, and also if the different survey types are analyzed separately in CH, CZ, IL and SI[5].

Table 14: Results for infit statistics and DIF-analyses for the HLS_{19}-Q12

Country	Item	Infit	*t*	*p*	Gender DIF		Age DIF		Education DIF		unordered response categories
					Magn.	*p*	Magn.	*p*	Magn.	*p*	
AT	4	1.029	0.93	1.000	0.15	<0.001[a,b]	0.13	<0.001			
	10	1.068	2.46	0.334					0.11	<0.001[a]	
	23	0.903	-3.35	0.023			0.10	<0.001			
	31	1.088	3.03	0.067					0.13	<0.001[b]	
	32	0.949	-1.77	1.000					0.14	<0.001[b]	X
	37	1.063	2.17	0.607			0.10	<0.001			
	42	0.976	-0.81	1.000							X[a]
BE	23	0.904	-2.31	0.612			0.21	<0.001[a]			
	31	1.177	3.94	0.003			0.17	<0.001			
CH_Total	16	0.973	-0.79	1.000							X
	23	0.975	-0.81	1.000			0.14	<0.001[b]			Random[f]
	31	1.150	5.13	<0.001							
	37	1.108	3.34	0.027			0.18	<0.001[d]			
	42	0.971	-0.98	1.000			0.10	<0.001			
CH_CAWI	16										X
	23						0.14	<0.001			
	31	1.153	5.07	<0.001							
	37						0.17	<0.001			
CH_CATI	10								0.42	<0.001	X
	16										X
	18										X
	23										X
	24										X

[5] The fact, that some tests for DIF are not significant for CATI in CH are partly due to the low sample size.

Country	Item	Infit	*t*	*p*	Gender DIF		Age DIF		Education DIF		unordered response categories
					Magn.	*p*	Magn.	*p*	Magn.	*p*	
	31										X
	32										X
	37										X
	44										X
CZ_Total	31	0.965	-0.92	1.000					0.12	<0.001	
	32	1.052	1.29	1.000			0.21	<0.001[a, b]	0.20	<0.001[b]	
CZ_CAWI	32						0.17	<0.001	0.18	<0.001	
CZ_CATI	4										X
	7										X
	10										X
	16										X
	18										X
	23										X
	24										X
	31										X
	32						0.39	<0.001	0.29	<0.001	X
	37										X
	42								0.23	<0.001	X
	44										X
DE	4	1.026	0.80	1.000	0.10	<0.001					X[a]
	23	0.940	-1.88	1.000							X[a]
	37	1.150	4.38	<0.001					0.11	<0.001[a]	X[a]
DK	7	1.036	1.46	1.000			0.09	<0.001			
	10	0.963	-1.60	1.000					0.07	<0.001	
	23	0.975	-1.04	1.000	0.11	<0.001[b]	0.17	<0.001[a, d]	0.10	<0.001[b]	
	32	0.896	-4.45	<0.001	0.12	<0.001[d]					
	37	1.073	2.86	0.103			0.09	<0.001			

Country	Item	Infit	*t*	*p*	Gender DIF		Age DIF		Education DIF		unordered response categories
					Magn.	*p*	Magn.	*p*	Magn.	*p*	
	42	1.034	1.39	1.000			0.21	<0.001[a, e]			
FR	4	1.171	4.88	<0.001[a]							X[a]
	7	1.016	0.50	1.000							X[a]
	16	1.062	1.815	1.000							X[a]
	23	0.942	-1.71	1.000			0.18	<0.001[a, b]			X[a]
	32	0.956	-1.25	1.000					0.11	<0.001[a]	
	37	1.114	3.07	0.055							X[a]
HU	4	1.190	3.84	0.004							X[a]
	16	0.939	-1.16	1.000							X[a]
	18	1.116	2.45	0.419							X[a]
IE	4	1.064	2.77	0.094	0.11	<0.001					X[a]
	7	1.049	230	0.301			0.15	<0.001[c]	0.12	<0.001[b]	X[a]
	10	1.057	2.75	0.095							X[a]
	16	0.944	-1.87	0.606	0.10	<0.001					Random[f]
	18	1.056	2.74	0.095			0.15	<0.001[a, d]			Random[f]
	23	0.904	-4.38	<0.001	0.10	<0.001					X[a]
	24	0.969	-1.32	0.928			0.10	<0.001			
	31	1.205	9.57	<0.001[a, d]					0.08	<0.001	X[a]
	32	0.955	-1.74	0.651							X[a]
	37	0.958	-1.88	0.606							X[a]
	42	1.037	1.61	0.748			0.21	<0.001[a, d]	0.09	<0.001	X[a]
	44	0.952	-2.09	0.441							X
IL_Total	16	0.984	-0.38	1.000							X[a]
	37	1.039	0.95	1.000							x
IL_CAWI	16										X
IL_CATI	4										X

Country	Item	Infit	*t*	*p*	Gender DIF		Age DIF		Education DIF		unordered response categories
					Magn.	*p*	Magn.	*p*	Magn.	*p*	
	7										X
	10										X
	16										X
	18										X
	23										X
	24										X
	31										X
	32										X
	37										X
	42										X
	44										X
NO	10	1.081	2.85	0.097							X[a]
	18	1.040	1.45	1.000							x
	31	1.162	5.60	<0.001					0.10	<0.001	x
	42	1.025	0.85	1.000			0.19	<0.001[c]			
PT	4	0.952	-0.72	1.000							X[a]
	7	1.014	0.22	1.000							X[a]
	10	1.032	0.63	1.000							X[a]
	16	0.970	-0.43	1.000							X[a]
	18	1.066	1.24	1.000							X[a]
	23	0.855	-2.33	0.590							X[a]
	24	0.863	-2.03	1.000							X[a]
	31	0.972	-0.49	1.000							X[a]
	32	0.806	-2.74	0.213							X[a]
	37	0.882	-1.49	1.000							X[a]
	42	0.906	-1.23	1.000							X[a]
	44	1.056	0.83	1.000							X[a]
RU	18	0.992	-0.37	1.000					0.09	<0.001	

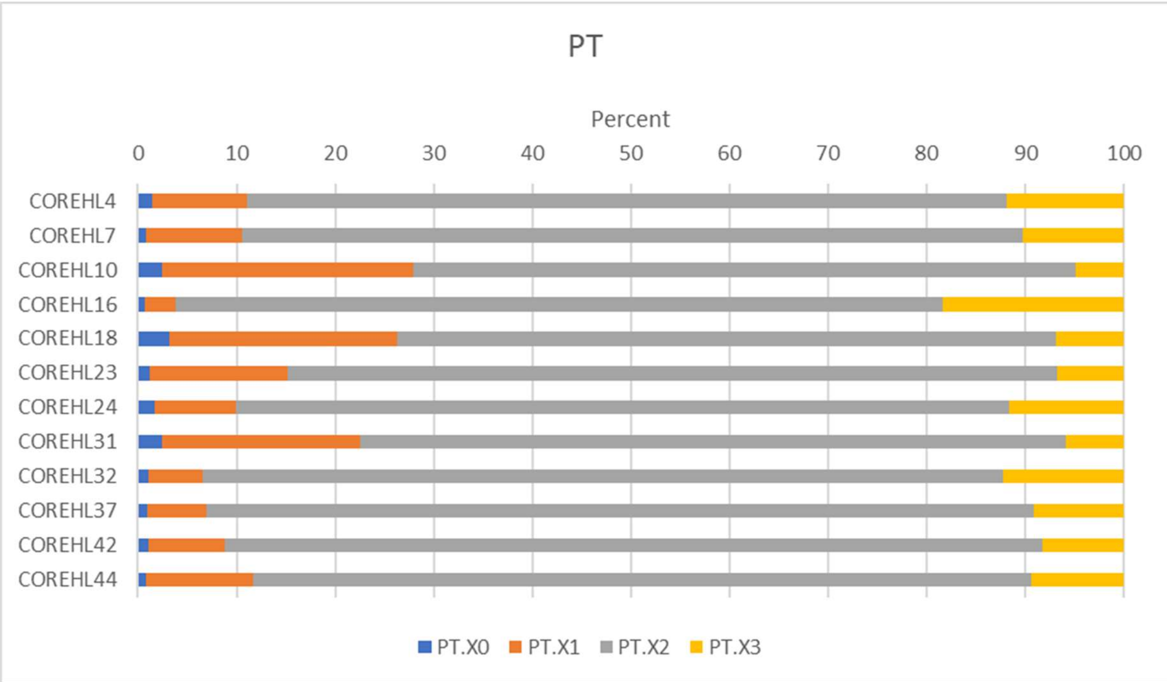
Country	Item	Infit	*t*	*p*	Gender DIF		Age DIF		Education DIF		unordered response categories
					Magn.	*p*	Magn.	*p*	Magn.	*p*	
	32	0.954	-1.90	0.860			0.06	<0.001			
	37	1.077	2.87	0.091			0.08	<0.001[b]			
SI_Total	10	1.073	2.76	0.070					0.10	<0.001	
	23	0.995	-0.19	1.000	0.14	<0.001[a]	0.21	<0.001[a, e]	0.14	<0.001[a, c]	
	24	0.840	-6.16	<0.001							
	31	1.350	12.69	<0.001[a, e]	0.09	<0.001	0.10	<0.001	0.21	<0.001[a, e]	
	32	0.893	-3.98	0.001			0.23	<0.001[a, e]	0.28	<0.001[a]	
	44	1.134	4.88	<0.001							
SI_CAWI	7						0.14	<0001			
	23				0.17	<0.001	0.45	<0.001			
	31	1.328	8.34	<0.001	0.12	<0.001			0.17	<.001	
	32						0.16	<0.001	0.22	<0.001	
	37										X
	42										X
	44	1.181	4.67	<0.001							
SI_CAPI	23				0.12	<0.001	0.18	<0.001			
	24	0.825	-4.65	<0.001							
	31	1.275	7.14	<0.001							
	32						0.17	<0.001	0.20	<0.001	
SK	32	0.886	-3.63	0.010					0.10	<0.001[e]	

Only significant results are shown in the table; [a]: significant also in the random sample of n=900; [b]: significant also in one of the four subsamples; [c]: significant also in two of the four subsamples; [d]: significant also in three of the four subsamples; [e]: significant also in all four subsamples; [f]: only in the random sample.

Considering the random sample, results are more favorable in most countries. In RU and SK none of the items displays significant DIF or significant misfit in the random sample.

Applying the Nominal Categories Model to identify the empirical ordering of the response categories revealed unordered response categories for all items in PT, for nine of the twelve items in IE, for five items in FR and for two or three items in AT, DE, HU, IL, and NO (for parameter estimates of the response categories see Table A7 in the Appendix). Closer inspection showed that the response category “very difficult” was very rarely chosen (< 1% of persons) in some items in several countries (see frequency distributions for the response categories in the different countries in Figure A1 in the Appendix). It seems that it was hard for the respondents to discriminate between the response options “very difficult” and “difficult”. On the other hand, in PT at least two third of the persons have chosen the answer ‘easy’ in all items (e.g. for Item 37 approximately 84%, see Figure 3). In the case of low endorsement rates in some of the categories the estimation of the parameters of the Nominal Categories Model can be affected such that response categories are tagged as unordered (cf. García-Pérez, 2018). Therefore, regarding the results for the survey mode CATI in CH, CZ and IL, the small sample sizes must be considered.

Figure 3: Frequency distributions for the response categories in PT



X0: very difficult; X1: difficult; X2: easy; X3: very easy

Dichotomous Rasch model:

Global model tests:

The global LR-test using the *split criterion median score* was not possible for AT, PT and SI, because the data were not well-conditioned (Fischer, 1981). Well-conditioned data means that in every possible partition of the items into two non-empty item subsets at least one person has chosen the answer category 1 on one item in the first subset and answer category 0 on one item in the other subset. This is a necessary and sufficient condition for a unique solution of the conditional maximum likelihood estimates of the item parameters (Fischer, 1981). Of the remaining twelve countries, four LR-tests were significant in the total samples and two of them also in the random sample ($p < 0.001$; see

Table 15).

Furthermore, the global LR-tests using the *split criterion median age* was significant in 11 countries in the total samples, and for 7 countries also in the random sample. In RU the data in the random sample were not well-conditioned, thus the LR-test could not be performed. Using the *split criterion gender* the LR tests were significant in 3 countries only in the total sample. For the *split criterion education* the LR tests were significant in 9 countries in the total sample only. Only for CZ and HU none of the LR tests was significant in the total samples and random samples, and in the random sample none of the LR tests was significant for AT, IL, SI and SK.

Table 15: Results of the LR-tests of the dichotomized HLS_{19}-Q12

Split Criterion	Country	Total Sample			Random Sample		
Median Score		**Chisq**	**DF**	**P**	**Chisq**	**DF**	**P**
	BE	16.67	11	0.118	17.30	11	0.099
	CH	**45.63**	**11**	**<0.001**	15.64	11	0.155
	CZ	18.65	11	0.068	12.08	11	0.358
	DE	**49.98**	**10**	**<0.001**	**40.78**	**10**	**<0.001**
	DK	**32.72**	**11**	**0.001**	8.67	11	0.652
	FR	29.12	11	0.002	27.36	11	0.004
	HU	20.41	11	0.040	18.38	11	0.073
	IE	**57.72**	**11**	**<0.001**	**35.83**	**11**	**<0.001**
	IL	21.08	11	0.033	19.54	11	0.052
	NO	24.12	11	0.012	11.10	11	0.435
	RU	**84.40**	**11**	**<0.001**	Not well-conditioned		
	SK	29.78	11	0.002	15.28	11	0.17
Age							
	AT	**38.74**	**11**	**<0.001**	23.05	11	0.017
	BE	**87.33**	**11**	**<0.001**	**69.41**	**11**	**<0.001**
	CH	**159.84**	**11**	**<0.001**	**50.96**	**11**	**<0.001**
	CZ	28.12	11	0.003	23.30	11	0.016
	DE	**33.18**	**11**	**<0.001**	18.29	12	0.075
	DK	**176.77**	**11**	**<0.001**	**51.78**	**11**	**<0.001**
	FR	**76.42**	**11**	**<0.001**	**33.72**	**11**	**<0.001**

Split Criterion	Country	Total Sample			Random Sample		
	HU	15.83	11	0.148	13.81	11	0.244
	IE	**115.56**	**11**	**<0.001**	26.73	11	0.005
	IL	20.66	11	0.037	20.17	11	0.043
	NO	**71.59**	**11**	**<0.001**	**38.86**	**11**	**<0.001**
	PT	**43.22**	**11**	**<0.001**	**44.67**	**11**	**<0.001**
	RU	**124.56**	**11**	**<0.001**	**35.79**	**11**	**<0.001**
	SI	**99.50**	**11**	**<0.001**	23.88	11	0.013
	SK	31.06	11	0.001	12.18	11	0.350
Gender							
	AT	28.55	11	0.003	19.56	11	0.052
	BE	21.02	11	0.033	19.20	11	0.058
	CH	**43.06**	**11**	**<0.001**	16.96	11	0.109
	CZ	28.45	11	0.003	10.65	11	0.473
	DE	25.33	11	0.008	12.71	11	0.313
	DK	**68.07**	**11**	**<0.001**	21.55	11	0.028
	FR	27.85	11	0.003	20.10	11	0.044
	HU	10.84	11	0.457	8.21	11	0.695
	IE	31.74	11	0.001	20.93	11	0.034
	IL	18.20	11	0.077	16.92	11	0.110
	NO	16.40	11	0.127	24.97	11	0.009
	PT	17.30	11	0.099	15.61	11	0.156
	RU	21.79	11	0.026	11.96	11	0.367
	SI	**35.56**	**11**	**<0.001**	16.17	11	0.135
	SK	15.99	11	0.141	7.58	11	0.750
Education							
	AT	**39.12**	**11**	**<0.001**	18.86	11	0.064
	BE	9.29	11	0.595	8.81	11	0.64
	CH	**42.11**	**11**	**<0.001**	28.03	11	0.003
	CZ	20.97	11	0.034	21.18	11	0.032

Split Criterion	Country	Total Sample			Random Sample		
	DE	**40.47**	**11**	**<0.001**	29.61	11	0.002
	DK	**44.07**	**11**	**<0.001**	23.01	11	0.018
	FR	17.39	11	0.097	15.50	11	0.161
	HU	22.74	11	0.019	22.30	11	0.022
	IE	32.08	11	0.001	13.75	11	0.247
	IL	**38.38**	**11**	**<0.001**	27.68	11	0.004
	NO	**38.37**	**11**	**<0.001**	15.68	11	0.153
	PT	27.83	11	0.003	26.27	11	0.006
	RU	**69.19**	**11**	**<0.001**	11.10	11	0.435
	SI	**78.07**	**11**	**<0.001**	27.68	11	0.004
	SK	**57.17**	**11**	**<0.001**	16.48	11	0.124

Only for AT and CZ the global model tests for local independence were not significant in the random samples (see Table 16: Results of the global model test for local independence in the total and the random samples).

Table 16: Results of the global model test for local independence in the total and the random samples

Country	*p* total sample	*p* random sample
AT	< 0.001	0.014
BE	< 0.001	< 0.001
CH	< 0.001	< 0.001
CZ	< 0.001	0.002
DE	< 0.001	< 0.001
DK	< 0.001	< 0.001
FR	< 0.001	< 0.001
HU	< 0.001	< 0.001
IE	n.a.	< 0.001
IL	< 0.001	< 0.001
NO	< 0.001	< 0.001
PT	< 0.001	< 0.001
RU	n.a.	< 0.001
SI	< 0.001	< 0.001
SK	< 0.001	< 0.001

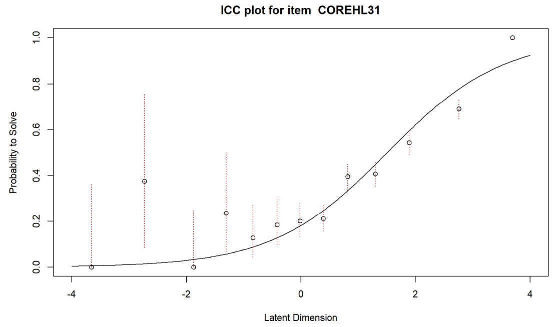
n.a.: statistic could not be calculated due to the large sample size

Analyses at item level:

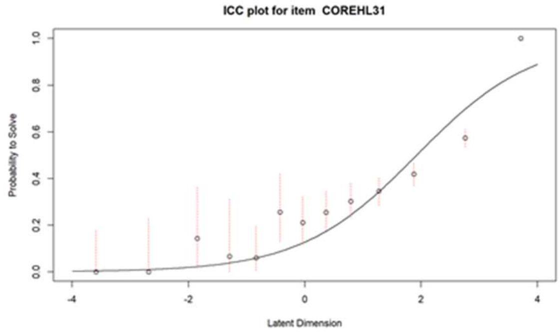
Item 31 again has infit values ≥ 1.10 with corresponding high t-values in several countries (CH, FR, NO, SI; cf. Table 17). Inspection of the ICC plots, for instance for NO and SI, reveals clear deviations of the observed scores from the expected values (cf. Figure 4).

Figure 4: ICC plots for Item 31 in NO and SI

NO



SI



The Fischer-Scheiblechner *z*-statistic (Wald test) was significant for several items and split criteria, therefore indicating DIF (cf. Table 17):

- Item 4 for
  - gender in AT, IE and
  - median test score in FR,
- Item 7 for age in IE, PT, RU,
- Item 10 for
  - gender in DK,
  - age in CH, DK, and
  - education in AT, IL,
- Item 16 for
  - age in RU,
- Item 18 for
  - gender in DK,
  - age in DK, IE, RU, and
  - education in RU,
- Item 23 for
  - gender in CH, DK, SI
  - age in BE, CH, DE, DK, FR, SI and
  - education in DK,
- Item 24 for median score in RU,
- Item 31 for
  - education in AT, NO, SI and
  - age in BE, DK, FR, NO
  - median test score in CH,
- Item 32 for
  - gender in DK,
  - age in NO, PT, RU, SI, and
  - education in IL, RU, SI,

- Item 37 for
  - gender in BE
  - age in AT, CH, DK, NO, RU
  - education in CH, DE, SK, and
  - median test score in DE,

- Item 42 for
    - age in CH, DK, IE, NO and
- Item 44 for education in SI.

Table 17: Results of item infit statistics and DIF analyses for the dichotomized HLS$_{19}$-Q12

Country	Item	Infit		Median DIF		Gender DIF		Age DIF		Education DIF	
		MSQ	t	z	P	z	P	z	P	Z	P
AT	4	0.90	-1.89			4.22	<0.001				
	10	1.03	1.60							3.57	<0.001[a]
	31	0.97	-1.11							3.40	0.001
	37	1.05	1.51					4.44	<0.001		
BE	23	0.94	-1.55					-4.80	<0.001[a]		
	31	1.08	2.33					5.06	<0.001[a]		
	37	1.04	0.87			-3.56	<0.001				
CH	10	1.01	0.25					3.63	<0.001		
	23	0.97	-0.95			-3.61	<0.001	-6.44	<0.001[a]		
	31	1.11	4.99	4.88	<0.001						
	37	1.09	2.71					7.97	<0.001[a]	-4.30	<0.001
	42	0.95	-1.73					-4.41	<0.001		
DE	23	0.91	-2.70					-4.28	<0.001		
	37	1.10	2.80	5.48	<0.001[a]						
DK	10	0.94	-3.32			3.51	<0.001	4.73[n]	<0.001	3.25	0.001
	18	0.97	-1.69			-3.88	<0.001	4.38	<0.001		
	23	0.89	-3.28			-3.36	0.001	-4.82	<0.001	-3.91	<0.001
	31	1.01	0.47					3.72	<0.001		
	32	0.83	-3.38			-3.52	<0.001				
	37	1.06	2.43					4.71	<0.001		
	42	0.96	-1.84					-8.58	<0.001		
	44	1.07	3.48	4.93	<0.001						
FR	4	1.08	2.00	4.03	<0.001[a]						
	23	0.92	-2.33					-5.58	<0.001		
	31	0.97	-1.16					4.71	<0.001		
IE	4	0.99	-0.46			4.03	<0.001				

Country	Item	Infit		Median DIF		Gender DIF		Age DIF		Education DIF	
		MSQ	t	z	P	z	P	z	P	Z	P
	7	1.02	0.88					5.10	<0.001		
	18	1.02	1.14					4.33	<0.001		
	31	1.14	8.72[a]	5.95	<0.001[a]						
	42	0.99	-0.47					-6.48	<0.001[a]		
IL	10	0.98	-0.48							3.51	<0.001
	32	0.92	-1.10							-4.87	<0.001[a]
NO	31	1.10	4.68	3.37	0.001			3.56	<0.001	3.71	<0.001
	32	0.92	-1.84					-3.25	0.001		
	37	0.96	-1.47					4.07	<0.001	3.30	0.001
	42	1.00	-0.02					-4.81	<0.001		
PT	7	1.13	1.46					-3.58	<0.001[a]		
	32	0.65	-3.38					3.71	<0.001[a]		
RU	7	0.99	-0.37					3.61	<0.001		
	16	0.94	-1.73					-3.67	<0.001		
	18	0.97	-1.33					5.41	<0.001	-4.68	<0.001
	24	0.90	-4.75	-3.66	<0.001						
	32	0.82	-5.69					4.14	<0.001	-4.00	<0.001
	37	1.08	2.76					-6.58	<0.001		
	44	1.08	4.09	6.95	<0.001						
SI	23	0.97	-0.64			-4.21	<0.001	-6.92[n]	<0.001		
	31	1.16	7.64[a]							5.49	<0.001
	32	0.79	-3.88					5.58	<0.001	-5.13	<0.001
	44	1.08	2.56							4.33	<0.001
SK	37	1.02	0.53	3.23	0.001					4.87	<0.001

[a]: also significant in the random sample of n=900; [n]: non-uniform DIF

The check of local independence using the *Q3*-statistic revealed dependent item pairs in all countries except for AT in the random sample (cf. Table 18).

Table 18: Number of item pairs with significant correlations of inter-item residuals in the different countries

Country	Total sample	Random sample
AT	5	0
BE	6	5
CH	5	1
CZ	6	2
DE	5	3
DK	10	3
FR	3	2
HU	5	4
IE	n.a.	4
IL	2	2
NO	6	3
PT	2	2
RU	n.a.	2
SI	4	2
SK	6	5

5. Summary

The short version of the HLS-EU questionnaire representing all 12 elements of the HL conceptual matrix (HLS$_{19}$-Q12) which was developed using data from the first wave of the HLS-EU study did not show acceptable fit to the PCM in all 15 countries. According to the PCA/t-test procedure, only in Norway the HLS$_{19}$-Q12 could be deemed sufficiently unidimensional. The values of the *SRMSR* statistic were above the cut-off value of 0.05 for good model fit suggested by Maydeu-Olivares (2013) in all countries except Israel, although below the less conservative value according to Hu and Bentler (1999). However, the global model test for local independence based on the adjusted *Q3*-statistic (Robitzsch et al., 2020) yielded significant results in all 15 countries both in the total samples as well as the random samples of n=900 drawn in each of the countries. Analyses on the item level revealed that for Russia and Slovakia none of the items displayed misfit, DIF or unordered response categories in the random samples of n=900. However, also in Russia and Slovakia – as in all other countries - some dependent item pairs were observed with residual correlations between r = 0.20 and r = 0.25 in the total or the random samples. In all other countries there was at least one item with unordered response categories, and/or at least one poor fitting item, and/or at least one item displaying DIF in the random sample, the total sample, and also in smaller subsamples. The most problematic items are 4, 23, 31, 32, 37 and 42, which showed significant model deviations in several countries. Items 18, 23, 31, 32, 37 and 42 already displayed DIF in previous analyses using the original HLS-EU-8 data (see 2.2). However, it needs to be considered that the response categories and also the wording of some items have been changed which of course could affect model fit either positively or negatively. The low endorsement rates in some of the categories could be the reason for the high number of items which were tagged as unordered in some countries (cf. García-Pérez, 2018). The frequency distribution for answer category “very difficult” varies from 0.30% (item 16) to 13.80% (item 7), for “difficult” from 3.15% (item 16) to 57.31% (item 10), for “easy” from 25.06% (item 10) to 83.95% (item 37), and for “very easy” from 3.82% (item 10) to 67.89% (item 16). However, the response category “very difficult” was very rarely chosen (< 1% of persons) in some items in several countries. Survey-type-specific analyses for CH, CZ, IL and SI revealed comparable results for CAWI and also for CAPI in SI. For CATI the results are somewhat different; however, sample sizes are (too) small, especially in CH (n=139) and IL (n=290).

For the dichotomous RM, model fit is barely acceptable for the Czech Republic and Austria. In both countries, in the random sample of n=900, neither the test of local dependency nor the LR tests were significant. However, in Austria item 10 displayed DIF for education both in the total and the random sample, and in the Czech Republic there are two dependent item pairs according to the *Q3* statistic. In all other countries the global test for local independence and / or at least one LR test was significant both in the total and the random samples. On item level, for the split criteria gender, age and education the results for the dichotomous scoring are quite similar to those for the polytomous scoring. Again, items 23, 31, 32, 37 and 42 display DIF in several countries and/or split criteria.

Exploring the most problematic items leads to following hypotheses:

- Item 23 displays DIF for age in many countries, whereby the item is relatively easier for the older age group. The examples given in the instructions could be the reason.
- Item 31 is generally a poor fitting item and displays DIF for education and age in several countries. This item is easier for people of the lower education group and for the younger age group. The reason could be that higher educated people and older people are more critical regarding information from the mass media. Therefore, a second dimension of critical appraisal of information provided by mass media might be responsible for the poor model fit of this item.

In those four countries, in which different survey methods were applied, analyses were conducted across the survey methods. Additional separate analyses for the different survey methods could be conducted. However, the overall conclusion that some items display significant misfit to both the PCM and the RM in several countries would not be affected.

6. References

Adams, R.J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*, 162-172.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267-281). Budapest, Hungary: Akademiai Kiado.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38,* 123- 140.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345-370.

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48(6),* 1-29. doi:10.18637/jss.v048.i06

Duong, T.V., Aringazina, A., Balsunova, G., Nurjanah Pham, T.V., Pham, K.M., Truong, T.Q., … Chang, P.W. (2017). Measuring health literacy in Asia: Validation of the HLS-EU-Q47 survey tool in six Asian countries. *Journal of Epidemiology, 27(2),* 80-86.

Finbråten, H.S., Pettersen, K.S., Wilde-Larsson, B., Norström, G., Trollvik, A., Guttersrud, Ø. (2017). Validating the European Health Literacy Survey Questionnaire in people with type 2 diabetes: Latent trait analyses applying multidimensional Rasch modelling and confirmatory factor analyis. *J Adv Nurs, 73,* 2730-2744.

Finbråten, H.S., Wilde-Larsson, B., Nordström, G., Pettersen, K.S., Trollvik, A. & Guttersrud, Ø. (2018). Establishing the HLS-Q12 short version of the European Health Literacy Survey Questionnaire: latent trait analyses applying Rasch modelling and confirmatory factor analysis. *BMC Health Services Research, 18,* 506. https://doi.org/10.1186/s12913-018-3275-7

Fischer, G.H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika, 46,* 59–77.

Fischer, G.H. & Scheiblechner, H. (1970). Algorithmen und Programme für das probabilistische Testmodell von Rasch. [Algorithms and programs for Rasch's probabilistic test model.] *Psychologische Beiträge, 12,* 23-51.

García-Pérez, M.A. (2018). Order-Constrained Estimation of Nominal Response Model Parameters to Assess the Empirical Order of Categories. *Educational and Psychological Measurement, 78(5),* 826-856.

Hagell, P. (2014). Testing Rating Scale Unidimensionality Using the Principal Component Analysis (PCA)/t-Test Protocol with the Rasch Model: The Primacy of Theory over Statistics. *Open Journal of Statistics, 4,* 456-465.

HLS-EU Consortium (2012): *Comparative report of health literacy in eight EU member states. The European Health Literacy Survey HLS-EU.* Online Publication: http:/www.health-literacy.eu

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika, 72,* 297-307.

Linacre, J.M. & Wright, B.D. (1994). *Rasch Measurement Transactions, Part 2, Volume 8(2).* https://www.rasch.org/rmt.

Mair, P., Hatzinger, R., Maier, M.J., Rusch, T., & Debelak, R. (2020). eRm: Extended Rasch Modeling. 1.0-1. http://cran.r-project.org/package=eRm

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

Maydeu-Olivares, A. (2013). Goodness-of-Fit Assessment of Item Response Theory Models. *Measurement, 11,* 71-101.

Pelikan, J.M. & Ganahl, K. (2017). Measuring Health Literacy in General Populations: Primary Findings from the HLS-EU Consortium's Health Literacy Assessment Effort. In R.A. Logan & E.R. Siegel (Eds.), *Health Literacy: New Directions in Research, Theory and Practice* (pp. 34-59). Amsterdam, Berlin, Washington, DC: IOS Press.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.* (Expanded edition. Chicago: The university of Chicago Press).

Robinson, M., Johnson, A.M., Walton, D.M., McDermid, J.C. (2019). A comparison of the polytomous Rasch analysis output of RUMM2030 and R (ltm/eRM/TAM/lordif). *BMC Medical Research Methodology, 19*:36.

Robitzsch, A. (2019). Package `sirt'. Version 3.3-26. March 18, 2019.

Robitzsch, A., Kiefer, T., & Wu, M. (2020). TAM: Test Analysis Modules. R package version 3.5-19. https://CRAN.R-project.org/package=TAM

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Smith Jr., E.V. (2002) Detecting and Evaluating the Impact of Multidimensionality Using Item Fit Statistics and Principal Component Analysis of Residuals. *Journal of Applied Measurement, 3,* 205-231.

Smith, A.B., Rush, R., Fallowfield, L.J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8:*33.

Sørensen, K., Van den Broucke, S., Pelikan, J.M., Fullam, J., Doyle, G., Slonska, Z., Kondilis, B., Stoffels, V, Osborne, R.H., Brand, H. (2013). Health literacy in populations: Illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). *BMC Public Health, 13(948).*

Sørensen, K., Pelikan, J.M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., Fullam, J., Kondilis, B., Agrafiotis, D., Uiters, E., Falcon, M., Mensing, M., Tchamov, K., van den Broucke, S., Brand, H. on behalf of the HLS-EU Consortium (2015). Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *European Journal of Public Health, 25(6),* 1053-1058.

Thissen, D., Cai, L., & Bock, R.D: (2010). The Nominal Categories Model. In M. L. Nering & R. Ostini (Ed.), *Handbook of Polytomous Item Response Theory Models* (pp. 43-75), Abingdon, Oxon: Routledge.

Weiss, B., Mays, M.Z., Martz, W., Castro, K.M., DeWalt, D.A., Pignone, M.P., Mockbee, J., Hale, F.A. (2005). Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med, 3(6),* 514-522.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125-145. doi: 10.1177/014662168400800201

Appendix

Table A1: Results of Step 1 of the development of the HLS-EU-Q12 in the total sample of the EU-8 countries

				DIF [magnitude]		
Item	Infit	*t*	p_Holm	Gender	Age	Education
1	1.027	1.175	1.000		o > y [0.196]	l > h [0.19]
2	0.983	-0.74	1.000		o > y [0.107]	l > h [0.112]
3	1.082	3.53	**0.028**			
4	1.009	0.358	1.000			
5	1.027	1.189	1.000		o < y [0.1]	
6	1.129	5.63	**0.000**		o > y [0.104]	l > h [0.092]
7	1.043	1.869	1.000			
8	0.926	-3.18	0.083			l > h [0.097]
9	0.931	-3.03	0.130			
10	1.008	0.39	1.000			
11	1.071	3.218	0.076		o < y [0.094]	l < h [0.113]
12	1.182	7.89	**0.000**			l < h [0.111]
13	0.912	-3.87	**0.008**			
14	1.023	0.925	1.000		o < y [0.105]	
15	1.16	5.81	**0.000**			
16	0.951	-2.04	1.000			
17	0.896	-4.5	**0.001**		o > y [0.076]	l > h [0.149]
18	0.939	-2.84	0.228		o > y [0.074]	
19	0.862	-6.43	**0.000**			
20	0.802	-9.22	**0.000**			l > h [0.095]
21	0.905	-4.01	**0.005**			l > h [0.115]
22	0.911	-3.92	**0.006**			l > h [0.075]
23	0.895	-4.57	**0.000**			l > h [0.079]
24	0.942	-2.46	0.662			
25	1.022	0.982	1.000		o < y [0.167]	l < h [0.141]
26	1	-0.01	1.000	f < m [0.083]	o < y [0.131]	
27	0.955	-2.05	1.000	f < m [0.091]	o < y [0.157]	
28	1.103	4.6	**0.000**			l < h [0.143]
29	1.143	6.2	**0.000**		o < y [0.196]	
30	1.122	5.08	**0.000**			l < h [0.227]
31	1.073	3.28	0.066			l > h [0.108]
32	0.816	-8.26	**0.000**		o > y [0.099]	l > h [0.173]
33	0.887	-5.11	**0.000**		o > y [0.085]	l < h [0.091]
34	1.121	5.4	**0.000**			l < h [0.095]
35	1.229	10	**0.000**			l < h [0.071]
36	1.018	0.778	1.000	f > m [0.093]		l > h [0.093]
37	1.014	0.557	1.000			l < h [0.114]
38	1.221	9.74	**0.000**		o > y [0.161]	
39	0.914	-3.83	**0.009**		o > y [0.104]	

40	0.855	-6.6	**0.000**	f < m [0.079]		
41	0.999	-0.06	1.000		o < y [0.119]	l < h [0.109]
42	0.982	-0.75	1.000		o < y [0.098]	l < h [0.098]
43	0.896	-4.36	**0.001**			
44	1.025	1.107	1.000			
45	1.284	11.5	**0.000**	f > m [0.086]	o > y [0.289]	l > h [0.207]
46	1.025	1.111	1.000			
47	1.179	7.81	**0.000**			

Table A2: Results of Step 3 of the development of the HLS-EU-Q12

Country	Item	infit	*t*	*p* Holm	DIF [magnitude]		
					Gender	Age	Education
AT	2	1.011	0.192	1		o > y [0.21]	
	4	1.047	0.765	1			
	5	1.014	0.254	1			
	7	1.021	0.374	1			
	8	0.959	-0.66	1			l > h [0.217]
	10	0.987	-0.2	1			
	11	1.096	1.641	1			l < h [0.191]
	13	0.934	-1.1	1			
	16	0.962	-0.61	1			
	18	0.876	-2.18	1			
	21	0.968	-0.48	1			
	23	0.973	-0.39	1			
	24	0.953	-0.77	1			
	28	1.113	1.915	1			
	31	1.085	1.448	1			l < h [0.203]
	33	0.898	-1.78	1			
	36	1.114	1.828	1			
	37	1.119	1.919	1			
	39	0.917	-1.43	1			
	42	1.001	0.027	1			
	43	0.94	-0.99	1			
	44	1.108	1.813	1			
BG	2	1.068	1.083	1			
	4	0.981	-0.26	1			
	5	0.949	-0.78	1			
	7	1.022	0.363	1			
	8	0.885	-1.86	1			
	10	1.035	0.566	1			
	11	1.176	2.667	0.459			
	13	0.925	-1.21	1			
	16	0.94	-0.89	1			
	18	1.036	0.583	1			
	21	0.942	-0.89	1			
	23	0.946	-0.81	1			
	24	0.921	-1.23	1			
	28	1.148	2.214	1			
	31	0.999	-0	1			
	33	1.037	0.605	1		o > y [0.286]	
	36	1.164	2.499	0.686			
	37	1.053	0.766	1			
	39	0.923	-1.23	1			
	42	1.008	0.135	1			
	43	0.901	-1.55	1			

	44	1.15	2.302	1			
					DIF [magnitude]		
Country	Item	infit	*t*	*p* Holm	Gender	Age	Education
EL	2	0.901	-1.67	1		o > y [0.208]	
	4	0.983	-0.24	1			
	5	1.019	0.309	1			
	7	1.109	1.713	1			
	8	1.053	0.8	1			
	10	1.034	0.588	1			
	11	1.011	0.205	1			
	13	0.823	-2.94	0.204			
	16	1.18	2.589	0.568			
	18	0.934	-1.08	1			
	21	0.901	-1.48	1			
	23	1.001	0.031	1		o < y [0.239]	
	24	1.015	0.246	1			
	28	1.202	3.208	0.084			
	31	1.109	1.751	1			
	33	0.882	-1.88	1		o > y [0.246]	
	36	1.002	0.052	1			
	37	1.035	0.518	1			
	39	0.86	-2.23	1		o > y [0.248]	
	42	0.937	-0.88	1			
	43	0.889	-1.67	1			
	44	1.206	3.194	0.087			
ES	2	1.071	1.001	1		o > y [0.221]	
	4	0.945	-0.69	1			
	5	0.987	-0.16	1			
	7	0.899	-1.42	1			
	8	0.998	-0	1			
	10	1.058	0.905	1			
	11	1.082	1.253	1			
	13	0.91	-1.41	1			
	16	1.036	0.579	1			
	18	1.162	2.398	1			
	21	0.894	-1.35	1			
	23	0.965	-0.53	1			
	24	0.955	-0.58	1			
	28	0.982	-0.26	1			
	31	1.092	1.461	1			
	33	1.023	0.315	1			
	36	1.018	0.282	1			
	37	0.962	-0.46	1			
	39	0.988	-0.17	1			
	42	0.969	-0.37	1			
	43	0.982	-0.23	1			
	44	1.067	0.876	1			

					DIF [magnitude]		
Country	Item	infit	*t*	*p* Holm	Gender	Age	Education
IE	2	1.134	1.816	1			
	4	1.145	1.846	1			
	5	0.99	-0.11	1			
	7	1.077	1.142	1			
	8	1.003	0.069	1			
	10	0.97	-0.46	1			
	11	1.057	0.913	1			
	13	0.986	-0.17	1			
	16	1.032	0.444	1			
	18	1.081	1.233	1			
	21	0.923	-1.06	1			
	23	0.93	-0.93	1			
	24	0.893	-1.5	1			
	28	1.22	3.242	0.075			
	31	1.065	1.018	1			
	33	0.897	-1.54	1			
	36	1.087	1.131	1			
	37	0.974	-0.36	1			
	39	0.994	-0.06	1			
	42	1.045	0.63	1			
	43	0.914	-1.2	1			
	44	0.946	-0.76	1			
NL	2	1.081	1.102	1		o > y [0.231]	
	4	1.039	0.514	1			
	5	1.037	0.553	1	f > m [0.233]		
	7	0.986	-0.19	1			
	8	0.909	-1.39	1			
	10	0.981	-0.29	1			
	11	1.082	1.278	1			
	13	1.026	0.399	1			
	16	0.906	-1.24	1			
	18	0.945	-0.8	1			
	21	0.892	-1.42	1			l > h [0.309]
	23	0.959	-0.53	1			
	24	0.981	-0.26	1			
	28	1.06	0.977	1			
	31	1.097	1.566	1			
	33	1.012	0.204	1			
	36	1.084	1.137	1			
	37	1.058	0.835	1			
	39	1.003	0.057	1			
	42	1.011	0.184	1		o < y [0.336]	
	43	0.936	-0.88	1	f < m [0.226]		
	44	1.176	2.638	0.467			

Country	Item	infit	*t*	*p* Holm	DIF [magnitude]		
					Gender	Age	Education
PL	2	1.017	0.279	1			
	4	1.154	1.997	1			
	5	1.17	2.261	1			
	7	0.908	-1.3	1			
	8	0.946	-0.74	1			
	10	1.047	0.721	1			
	11	1.085	1.23	1			
	13	0.866	-1.92	1			
	16	0.972	-0.39	1			
	18	1.011	0.188	1			
	21	0.853	-2.13	1			
	23	0.859	-1.98	1			
	24	0.961	-0.52	1			
	28	0.968	-0.46	1			
	31	1.057	0.852	1			
	33	0.947	-0.75	1			
	36	1.132	1.745	1			
	37	1.095	1.242	1			
	39	1.033	0.451	1			
	42	1.033	0.472	1			
	43	0.936	-0.89	1			
	44	1.016	0.244	1			
DE	2	1.05	0.814	1			
	4	1.021	0.336	1			
	5	0.999	-0.01	1			
	7	1.079	1.283	1			
	8	0.876	-1.8	1			l > h [0.234]
	10	1.083	1.362	1			
	11	1.052	0.868	1			
	13	0.902	-1.65	1			
	16	0.877	-1.79	1			
	18	1.038	0.627	1			
	21	0.826	-2.71	0.372			
	23	0.839	-2.59	0.493			
	24	1.005	0.099	1			
	28	1.226	3.473	**0.032**			
	31	1.108	1.77	1			
	33	0.956	-0.7	1			
	36	1.459	5.751	**0.000**			
	37	0.89	-1.83	1			
	39	0.901	-1.64	1			
	42	1.175	2.691	0.384			
	43	0.864	-2.26	1			
	44	1.074	1.221	1			

Table A3: Results of Step 4 of the development of the HLS-EU-Q12

Country	Item	infit	t	pHolm	DIF [magnitude]		
					Gender	Age	Education
AT	4	1	0.576	1			
	7	1.1	1.105	1			
	10	1	0.386	1			
	16	1	-0.72	1			
	18	0.9	-2.445	0.493			
	23	1	-0.582	1			
	24	0.9	-1.389	1			
	31	1.1	2.035	1			l < h [0.195]
	33	0.9	-2.151	0.975			
	37	1.1	1.904	1			
	42	1	0.276	1			
	44	1.1	1.53	1			
BG	4	1	-0.059	1			
	7	1	0.102	1			
	10	1.1	1.383	1			
	16	0.9	-0.796	1			
	18	1	0.726	1			
	23	0.9	-1.235	1			
	24	0.9	-1.163	1			
	31	1	-0.335	1			
	33	1	0.405	1		o > y [0.284]	l > h [0.196]
	37	1	0.251	1			
	42	1	0.601	1			
	44	1.1	1.66	1			
EL	4	1	-0.422	1			
	7	1.1	1.255	1			
	10	1	0.593	1			
	16	1.1	1.9	1			
	18	1	-0.698	1			
	23	0.9	-0.984	1			
	24	1	-0.291	1			
	31	1.1	1.272	1			
	33	0.9	-2.041	1		o > y [0.276]	
	37	1	0.164	1			
	42	0.9	-1.146	1			
	44	1.1	1.668	1			

Country	Item	infit	t	pHolm	DIF [magnitude]		
					Gender	Age	Education
ES	4	1	-0.36	1			
	7	0.9	-1.236	1			
	10	1.1	1.646	1			
	16	1	-0.003	1			
	18	1.1	1.995	1			
	23	0.9	-1.269	1			
	24	0.9	-0.96	1			
	31	1.1	1.487	1			
	33	1	0.256	1			
	37	0.9	-0.704	1			
	42	1	-0.449	1			
	44	1	0.349	1			
IE	4	1.2	2.017	1			
	7	1.1	1.398	1			
	10	1	0.363	1			
	16	1.1	0.747	1			
	18	1.1	1.146	1			
	23	0.9	-1.067	1			
	24	0.9	-1.604	1			
	31	1.1	1.516	1			
	33	0.9	-1.157	1			
	37	1	-0.353	1			
	42	1.1	0.793	1			
	44	1	-0.593	1			
NL	4	1	0.468	1			
	7	1	-0.215	1			
	10	1	0.337	1			
	16	0.9	-1.038	1			
	18	1	-0.45	1			
	23	1	-0.634	1			
	24	1	-0.245	1			
	31	1.1	1.063	1			
	33	1	0.23	1			
	37	1.1	1.134	1			
	42	1	-0.283	1		o < y [0.312]	
	44	1.2	2.453	0.496			

Country	Item	infit	t	pHolm	DIF [magnitude]		
					Gender	Age	Education
PL	4	1.1	1.666	1			
	7	1	-0.484	1			
	10	1.1	0.904	1			
	16	1	-0.461	1			
	18	1	-0.182	1			
	23	0.9	-1.973	1			
	24	1	-0.585	1			
	31	1.1	0.78	1			
	33	0.9	-1.187	1			
	37	1.1	1.376	1			
	42	1	0.661	1			
	44	1	-0.042	1			
DE	4	1.1	1.362	1			
	7	1.1	1.01	1			
	10	1.1	1.832	1			
	16	0.9	-0.823	1			
	18	1	0.24	1			
	23	0.8	-2.493	0.367			
	24	1	-0.33	1			
	31	1.1	1.507	1			
	33	0.9	-1.056	1			
	37	0.9	-1.778	1			
	42	1.2	2.513	0.359			
	44	1	0.777	1			

Table A4: Deviance and information criteria for the different questionnaire versions

	HLS-EU-Q47	HLS-EU-Q16	HLS-EU-Q12	HLS-Q12	HL-SF12
Deviance					
AT (n=537)	49611.48	16967.97	13130.16	13392.91	13078.13
BG (n=509)	43225.61	14486.94	11309.28	11773.31	12191.50
EL (n=517)	44260.85	14889.04	11713.98	11988.47	12226.51
ES (n=490)	36357.60	12388.97	9370.56	9605.61	10237.56
IE (n=492)	40568.56	14194.64	10756.19	11111.85	11140.04
NL (n=498)	40929.14	13574.60	10749.74	10963.51	10759.08
PL (n=493)	36736.61	12738.11	9835.49	9911.04	10449.65
DE (n=512)	45168.99	15246.92	12001.58	12457.90	12036.25
AIC					
AT	49895.48	17065.967	13204.16	13466.91	13152.13
BG	43509.61	14584.94	11383.28	11847.31	12265.50
EL	44544.85	14987.042	11787.98	12062.47	12300.51
ES	36641.6	12486.973	9444.56	9679.61	10311.56
IE	40852.56	14292.642	10830.19	11185.85	11214.04
NL	41213.14	13672.603	10823.74	11037.51	10833.08
PL	37020.61	12836.113	9909.49	9985.04	10523.65
DE	45452.99	15344.922	12075.58	12531.90	12110.245
CAIC					
AT	50646.09	17324.98	13399.74	13662.49	13347.71
BG	44252.62	14841.33	11576.88	12040.91	12459.11
EL	45290.07	15244.20	11982.16	12256.65	12494.69
ES	37379.21	12741.50	9636.75	9871.80	10503.75
IE	41590.74	14547.37	11022.54	11378.19	11406.38
NL	41953.05	13927.92	11016.53	11230.30	11025.87
PL	37759.08	13090.94	10101.91	10177.45	10716.07
DE	46196.83	15601.60	12269.40	12725.72	12304.07
AICc					
AT	49998.55	17076.03	13209.79	13472.55	13157.77
BG	43620.57	14595.62	11389.25	11853.28	12271.47
EL	44653.44	14997.53	11793.86	12068.34	12306.38
ES	36758.64	12498.11	9450.78	9685.832	10317.78
IE	40968.93	14303.73	10836.39	11192.04	11220.23
NL	41327.54	13683.54	10829.85	11043.62	10839.20
PL	37136.64	12847.17	9915.67	9991.22	10529.83
DE	45563.05	15355.53	12081.51	12537.83	12116.18
BIC					
AT	50504.09	17275.98	13362.74	13625.49	13310.71
BG	44110.62	14792.33	11539.88	12003.91	12422.11
EL	45148.07	15195.20	11945.16	12219.65	12457.69
ES	37237.21	12692.50	9599.75	9834.80	10466.75
IE	41448.74	14498.37	10985.54	11341.19	11369.38
NL	41811.05	13878.92	10979.53	11193.30	10988.87
PL	37617.08	13041.94	10064.91	10140.45	10679.07
DE	46054.83	15552.6	12232.40	12688.72	12267.07
No. of estimated parameters	142	49	37	37	37

Table A5: Results of Step 4 for the HLS-EU-Q12$_{32}$

Country	Item	infit	*t*	*p* Holm	DIF [magnitude]		
					Gender	Age	Education
AT	4	1.048	0.783	1			
	7	1.062	1.074	1			
	10	1.04	0.7	1			
	16	0.93	-1.15	1			
	18	0.876	-2.176	0.975			
	23	0.946	-0.821	1			
	24	0.911	-1.51	1			
	31	1.136	2.273	0.784			l < h [0.2]
	32	0.898	-1.739	1			
	37	1.088	1.439	1			
	42	1.011	0.201	1			
	44	1.079	1.356	1			
BG	4	0.973	-0.397	1			
	7	1.024	0.389	1			
	10	1.089	1.408	1			
	16	0.934	-0.984	1			
	18	1.073	1.158	1			
	23	0.91	-1.381	1			
	24	0.91	-1.417	1			
	31	0.984	-0.245	1		o > y [0.185]	
	32	1.008	0.139	1		o > y [0.252]	l > h [0.292]
	37	1.007	0.118	1			
	42	1.04	0.593	1			
	44	1.13	2.021	1			
EL	4	0.977	-0.331	1			
	7	1.081	1.3	1			
	10	1.043	0.739	1			
	16	1.127	1.849	1			
	18	0.961	-0.639	1			
	23	0.934	-0.983	1			
	24	0.979	-0.299	1			
	31	1.083	1.359	1			
	32	0.881	-1.848	1		o > y [0.189]	
	37	1.001	0.038	1			
	42	0.893	-1.546	1			
	44	1.095	1.537	1			

Country	Item	infit	*t*	*p* Holm	DIF [magnitude]		
					Gender	Age	Education
ES	4	0.961	-0.491	1			
	7	0.917	-1.18	1			
	10	1.119	1.813	1			
	16	0.999	-0.004	1			
	18	1.165	2.435	0.536			
	23	0.942	-0.931	1			
	24	0.919	-1.114	1			
	31	1.092	1.457	1			
	32	0.958	-0.518	1			
	37	0.946	-0.69	1			
	42	0.977	-0.271	1			
	44	1.03	0.421	1			
IE	4	1.148	1.889	1			
	7	1.092	1.351	1			
	10	1.046	0.72	1			
	16	1.03	0.422	1			
	18	1.144	2.151	0.881			
	23	0.922	-1.053	1			
	24	0.871	-1.841	1			
	31	1.115	1.76	1			
	32	0.875	-1.762	1			l > h [0.238]
	37	0.976	-0.336	1			
	42	1.077	1.063	1			
	44	0.952	-0.667	1			
NL	4	1.04	0.536	1			
	7	0.975	-0.387	1			
	10	1.032	0.524	1			
	16	0.923	-1.024	1			
	18	1.001	0.032	1			
	23	0.957	-0.561	1			
	24	0.982	-0.242	1			
	31	1.064	1.063	1			
	32	0.937	-0.757	1			
	37	1.072	1.041	1			
	42	0.987	-0.185	1		o < y [0.316]	
	44	1.17	2.577	0.349			

Country	Item	infit	*t*	*p* Holm	DIF [magnitude]		
					Gender	Age	Education
PL	4	1.127	1.691	1			
	7	0.978	-0.287	1			
	10	1.086	1.291	1			
	16	0.959	-0.589	1			
	18	1.018	0.285	1			
	23	0.861	-1.967	1			
	24	0.948	-0.714	1			
	31	1.052	0.774	1			
	32	0.847	-2.311	0.624			
	37	1.091	1.208	1			
	42	1.056	0.798	1			
	44	1.013	0.201	1			
DE	4	1.093	1.38	1			
	7	1.062	1.013	1			
	10	1.134	2.158	0.742			
	16	0.927	-1.028	1			
	18	1.04	0.671	1			
	23	0.841	-2.566	0.278			
	24	0.98	-0.284	1			
	31	1.114	1.876	1			
	32	0.833	-2.769	0.163			
	37	0.888	-1.873	1			
	42	1.179	2.76	0.163			
	44	1.057	0.959	1			

Table A6: Infit statistics for single items in the total samples of the different countries

Country	Item	Infit	*t*	*p*
AT				
	4	1.029	0.926	1.000
	7	1.008	0.279	1.000
	10	1.068	2.459	0.334
	16	1.012	0.420	1.000
	18	1.015	0.562	1.000
	23	0.903	-3.346	0.023
	24	1.011	0.393	1.000
	31	1.088	3.027	0.067
	32	0.949	-1.767	1.000
	37	1.063	2.166	0.607
	42	0.976	-0.812	1.000
	44	0.959	-1.467	1.000
BE				
	4	1.018	0.405	1.000
	7	1.059	1.329	1.000
	10	0.975	-0.577	1.000
	16	0.970	-0.592	1.000
	18	1.073	1.666	1.000
	23	0.904	-2.306	0.612
	24	0.968	-0.729	1.000
	31	1.177	3.938	0.003
	32	0.907	-2.054	1.000
	37	1.054	1.180	1.000
	42	0.948	-1.213	1.000
	44	0.935	-1.526	1.000
CH				
	4	1.065	2.103	0.688
	7	0.941	-1.928	0.862
	10	1.001	0.043	1.000
	16	0.973	-0.789	1.000

Country	Item	Infit	*t*	*p*
	18	0.940	-2.185	0.635
	23	0.975	-0.808	1.000
	24	1.015	0.504	1.000
	31	1.150	5.126	<0.001
	32	0.963	-1.205	1.000
	37	1.108	3.343	0.027
	42	0.971	-0.979	1.000
	44	0.954	-1.628	1.000
CZ				
	4	0.925	-1.862	1.000
	7	0.929	-1.788	1.000
	10	0.997	-0.078	1.000
	16	1.041	1.020	1.000
	18	0.994	-0.150	1.000
	23	0.985	-0.363	1.000
	24	1.044	1.130	1.000
	31	0.965	-0.921	1.000
	32	1.052	1.287	1.000
	37	1.066	1.571	1.000
	42	0.966	-0.880	1.000
	44	1.124	3.278	0.035
DE				
	4	1.026	0.795	1.000
	7	0.961	-1.278	1.000
	10	0.949	-1.608	1.000
	16	1.007	0.223	1.000
	18	0.971	-0.958	1.000
	23	0.940	-1.877	1.000
	24	0.977	-0.758	1.000
	31	1.012	0.386	1.000
	32	0.945	-1.785	1.000
	37	1.150	4.384	<0.001

Country	Item	Infit	*t*	*p*
	42	1.039	1.319	1.000
	44	1.037	1.192	1.000
DK				
	4	1.003	0.121	1.000
	7	1.036	1.459	1.000
	10	0.963	-1.602	1.000
	16	0.926	-3.002	0.070
	18	1.021	0.893	1.000
	23	0.975	-1.042	1.000
	24	0.974	-1.066	1.000
	31	1.048	2.035	0.838
	32	0.896	-4.454	<0.001
	37	1.073	2.855	0.103
	42	1.034	1.393	1.000
	44	1.096	4.052	0.002
FR				
	4	1.171	4.878	<0.001
	7	1.016	0.495	1.000
	10	1.020	0.660	1.000
	16	1.062	1.815	1.000
	18	0.964	-1.149	1.000
	23	0.942	-1.713	1.000
	24	0.920	-2.388	0.390
	31	1.046	1.472	1.000
	32	0.956	-1.251	1.000
	37	1.114	3.071	0.055
	42	0.929	-2.220	0.554
	44	0.945	-1.758	1.000
HU				
	4	1.190	3.838	0.004
	7	1.001	0.027	1.000
	10	0.957	-0.972	1.000

Country	Item	Infit	*t*	*p*
	16	0.939	-1.162	1.000
	18	1.116	2.446	0.419
	23	0.905	-1.815	1.000
	24	0.940	-1.189	1.000
	31	1.164	3.647	0.009
	32	0.894	-2.018	1.000
	37	0.950	-0.911	1.000
	42	0.938	-1.156	1.000
	44	0.984	-0.336	1.000
IE				
	4	1.064	2.773	0.094
	7	1.049	2.299	0.301
	10	1.057	2.752	0.095
	16	0.944	-1.869	0.606
	18	1.056	2.742	0.095
	23	0.904	-4.375	<0.001
	24	0.969	-1.323	0.928
	31	1.205	9.567	<0.001
	32	0.955	-1.743	0.651
	37	0.958	-1.876	0.606
	42	1.037	1.612	0.748
	44	0.952	-2.089	0.441
IL				
	4	1.082	1.903	1.000
	7	1.009	0.244	1.000
	10	1.003	0.090	1.000
	16	0.984	-0.380	1.000
	18	1.002	0.052	1.000
	23	0.931	-1.763	1.000
	24	1.050	1.249	1.000
	31	0.980	-0.519	1.000
	32	1.016	0.389	1.000

Country	Item	Infit	*t*	*p*
	37	1.039	0.954	1.000
	42	1.021	0.544	1.000
	44	1.018	0.465	1.000
NO				
	4	1.046	1.441	1.000
	7	1.025	0.895	1.000
	10	1.081	2.847	0.097
	16	0.902	-3.123	0.043
	18	1.040	1.448	1.000
	23	0.894	-3.716	0.005
	24	0.945	-1.955	0.708
	31	1.162	5.603	<0.001
	32	0.963	-1.223	1.000
	37	0.970	-1.033	1.000
	42	1.025	0.845	1.000
	44	1.020	0.725	1.000
PT				
	4	0.952	-0.715	1.000
	7	1.014	0.215	1.000
	10	1.032	0.629	1.000
	16	0.970	-0.427	1.000
	18	1.066	1.237	1.000
	23	0.855	-2.328	0.590
	24	0.863	-2.033	1.000
	31	0.972	-0.487	1.000
	32	0.806	-2.743	0.213
	37	0.882	-1.489	1.000
	42	0.906	-1.230	1.000
	44	1.056	0.834	1.000
RU				
	4	1.035	1.338	1.000
	7	0.961	-1.771	0.966

Country	Item	Infit	*t*	*p*
	10	0.961	-1.943	0.832
	16	0.976	-0.900	1.000
	18	0.992	-0.365	1.000
	23	0.932	-2.618	0.168
	24	0.935	-2.868	0.091
	31	0.969	-1.420	1.000
	32	0.954	-1.901	0.860
	37	1.077	2.867	0.091
	42	0.977	-0.980	1.000
	44	1.088	3.860	0.003
SI				
	4	1.008	0.294	1.000
	7	0.896	-3.809	0.002
	10	1.073	2.756	0.070
	16	0.908	-3.448	0.008
	18	0.915	-3.277	0.014
	23	0.995	-0.187	1.000
	24	0.840	-6.158	<0.001
	31	1.350	12.690	<0.001
	32	0.893	-3.975	0.001
	37	0.991	-0.316	1.000
	42	0.925	-2.705	0.075
	44	1.134	4.875	<0.001
SK				
	4	0.988	-0.369	1.000
	7	0.963	-1.176	1.000
	10	0.997	-0.107	1.000
	16	1.053	1.570	1.000
	18	0.945	-1.889	1.000
	23	0.952	-1.556	1.000
	24	1.009	0.300	1.000
	31	1.078	2.559	0.315

Country	Item	Infit	*t*	*p*
	32	0.886	-3.628	0.010
	37	1.067	1.975	1.000
	42	1.005	0.166	1.000
	44	1.092	3.032	0.078

Table A7: Parameter of the answer categories in the total samples of the different countries

Country	Item	Very Difficult	Difficult	Easy	Very Easy
AT					
	4	0	0.070	0.927	3
	7	0	0.565	1.262	3
	10	0	0.216	1.084	3
	16	0	0.963	1.423	3
	18	0	0.400	0.960	3
	23	0	0.236	0.892	3
	24	0	0.553	1.130	3
	31	0	0.058	0.765	3
	32	0	-0.970	0.124	3
	37	0	0.148	0.552	3
	42	0	-0.280	0.527	3
	44	0	0.094	0.744	3
BE					
	4	0	1.043	2.032	3
	7	0	0.875	1.884	3
	10	0	0.639	1.727	3
	16	0	0.988	1.885	3
	18	0	0.735	1.851	3

Country	Item	Very Difficult	Difficult	Easy	Very Easy
	23	0	0.745	1.581	3
	24	0	0.963	2.006	3
	31	0	0.368	1.445	3
	32	0	1.196	2.110	3
	37	0	0.836	1.756	3
	42	0	0.768	1.628	3
	44	0	0.694	1.520	3
CH					
	4	0	0.518	1.576	3
	7	0	0.646	1.531	3
	10	0	0.404	1.132	3
	16	0	-0.588	0.721	3
	18	0	0.452	1.263	3
	23	0	0.272	1.084	3
	24	0	0.199	1.368	3
	31	0	0.108	0.875	3
	32	0	0.687	1.618	3
	37	0	0.278	0.910	3
	42	0	0.458	1.273	3
	44	0	0.489	1.281	3
CZ					
	4	0	0.499	1.482	3
	7	0	0.617	1.534	3
	10	0	0.486	1.261	3
	16	0	0.594	1.391	3
	18	0	0.755	1.493	3

Country	Item	Very Difficult	Difficult	Easy	Very Easy
	23	0	0.808	1.675	3
	24	0	0.755	1.756	3
	31	0	0.349	1.321	3
	32	0	0.361	1.662	3
	37	0	0.562	1.389	3
	42	0	0.457	1.556	3
	44	0	0.463	1.358	3
DE					
	4	0	-0.471	1.052	3
	7	0	0.565	1.549	3
	10	0	0.563	1.349	3
	16	0	0.090	1.402	3
	18	0	0.556	1.570	3
	23	0	-0.099	1.134	3
	24	0	0.411	1.604	3
	31	0	0.485	1.421	3
	32	0	0.463	1.825	3
	37	0	-0.270	0.647	3
	42	0	0.479	1.419	3
	44	0	0.289	1.328	3
DK					
	4	0	0.499	1.469	3
	7	0	0.926	1.744	3
	10	0	0.623	1.366	3
	16	0	0.536	1.387	3
	18	0	0.431	1.167	3

Country	Item	Very Difficult	Difficult	Easy	Very Easy
	23	0	0.562	1.434	3
	24	0	0.765	1.522	3
	31	0	0.574	1.355	3
	32	0	0.746	1.566	3
	37	0	0.372	0.935	3
	42	0	0.599	1.257	3
	44	0	0.759	1.314	3
FR					
	4	0	-0.507	0.661	3
	7	0	-0.342	0.811	3
	10	0	0.195	1.021	3
	16	0	-2.912	-0.783	3
	18	0	0.237	1.129	3
	23	0	-0.116	0.846	3
	24	0	0.182	1.145	3
	31	0	0.224	0.967	3
	32	0	0.164	1.320	3
	37	0	-0.210	0.657	3
	42	0	0.093	0.972	3
	44	0	0.143	0.946	3
HU					
	4	0	-0.077	1.154	3
	7	0	0.194	1.178	3
	10	0	0.382	1.015	3
	16	0	-0.219	0.921	3
	18	0	-0.282	0.377	3

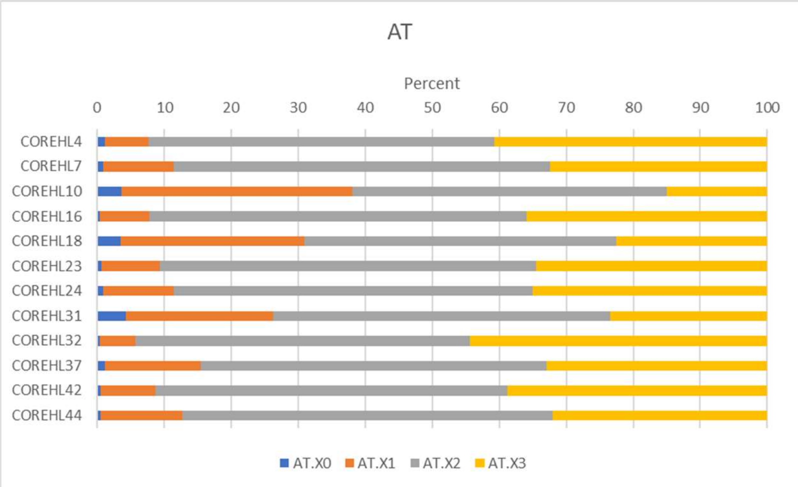
Country	Item	Very Difficult	Difficult	Easy	Very Easy
	23	0	0.057	1.222	3
	24	0	0.605	1.365	3
	31	0	0.192	0.689	3
	32	0	0.483	1.594	3
	37	0	0.702	1.413	3
	42	0	0.445	1.469	3
	44	0	0.385	0.945	3
IE					
	4	0	-0.423	0.352	3
	7	0	-0.533	0.336	3
	10	0	-0.047	1.144	3
	16	0	0.679	0.690	3
	18	0	0.277	0.976	3
	23	0	-0.239	0.601	3
	24	0	0.169	0.894	3
	31	0	-0.175	0.729	3
	32	0	-0.953	-0.076	3
	37	0	-0.309	0.307	3
	42	0	-0.387	0.372	3
	44	0	-0.129	0.440	3
IL					
	4	0	0.061	1.092	3
	7	0	0.209	1.082	3
	10	0	0.376	1.195	3
	16	0	-0.643	0.482	3
	18	0	0.254	1.022	3

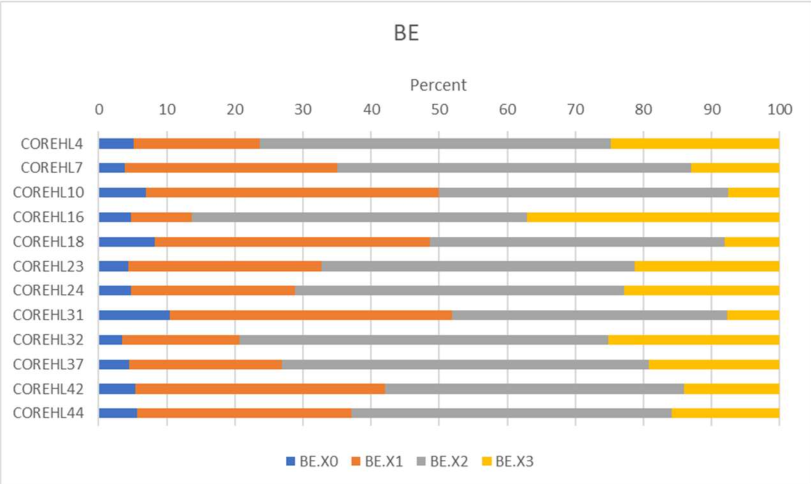
Country	Item	Very Difficult	Difficult	Easy	Very Easy
	23	0	0.022	0.791	3
	24	0	0.101	0.840	3
	31	0	0.135	0.788	3
	32	0	0.476	1.308	3
	37	0	-0.084	0.801	3
	42	0	0.200	0.710	3
	44	0	0.132	0.735	3
NO					
	4	0	0.417	0.906	3
	7	0	0.412	1.204	3
	10	0	-0.299	0.828	3
	16	0	0.714	1.411	3
	18	0	-0.056	0.853	3
	23	0	0.660	1.313	3
	24	0	0.482	1.262	3
	31	0	-0.059	0.683	3
	32	0	0.576	1.026	3
	37	0	0.323	0.984	3
	42	0	0.168	0.810	3
	44	0	0.296	0.887	3
PT					
	4	0	-9.311	-22.470	3
	7	0	-7.554	-16.627	3
	10	0	-0.922	-1.946	3
	16	0	-3.440	-33.517	3
	18	0	-2.439	-4.979	3

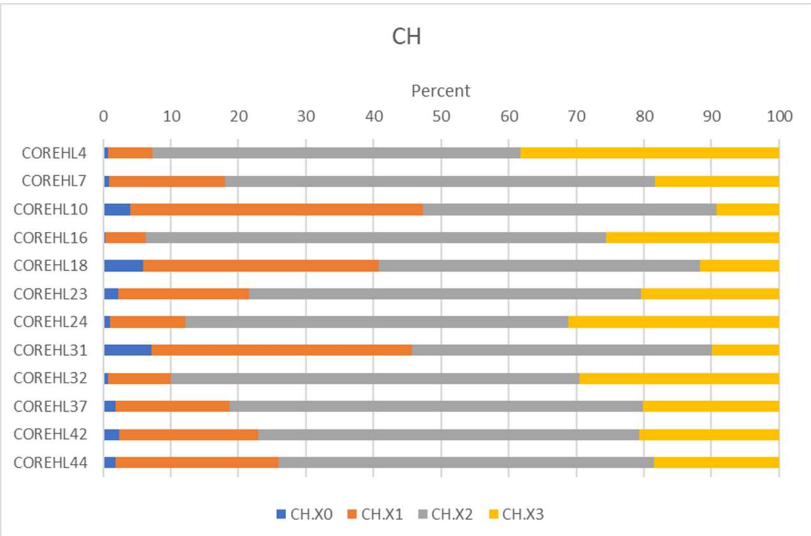
Country	Item	Very Difficult	Difficult	Easy	Very Easy
	23	0	-4.092	-7.742	3
	24	0	-10.045	-20.545	3
	31	0	-1.229	-2.786	3
	32	0	-8.535	-29.424	3
	37	0	-8.274	-16.101	3
	42	0	-4.084	-11.217	3
	44	0	-10.745	-19.270	3
RU					
	4	0	0.514	1.544	3
	7	0	0.702	1.555	3
	10	0	0.642	1.438	3
	16	0	0.697	1.667	3
	18	0	0.536	1.480	3
	23	0	0.553	1.610	3
	24	0	0.626	1.639	3
	31	0	0.407	1.310	3
	32	0	0.624	1.899	3
	37	0	0.437	1.330	3
	42	0	0.490	1.438	3
	44	0	0.104	1.101	3
SI					
	4	0	0.644	1.436	3
	7	0	0.506	1.393	3
	10	0	0.733	1.389	3
	16	0	0.349	1.389	3
	18	0	0.459	1.152	3

Country	Item	Very Difficult	Difficult	Easy	Very Easy
	23	0	0.512	1.171	3
	24	0	0.703	1.424	3
	31	0	0.362	0.800	3
	32	0	0.695	1.643	3
	37	0	0.444	1.098	3
	42	0	0.523	1.345	3
	44	0	0.682	1.159	3
SK					
	4	0	0.635	1.648	3
	7	0	0.935	1.878	3
	10	0	0.771	1.571	3
	16	0	0.612	1.445	3
	18	0	0.730	1.694	3
	23	0	0.746	1.493	3
	24	0	0.620	1.529	3
	31	0	0.514	1.290	3
	32	0	0.584	1.730	3
	37	0	0.700	1.416	3
	42	0	0.867	1.550	3
	44	0	0.576	1.475	3

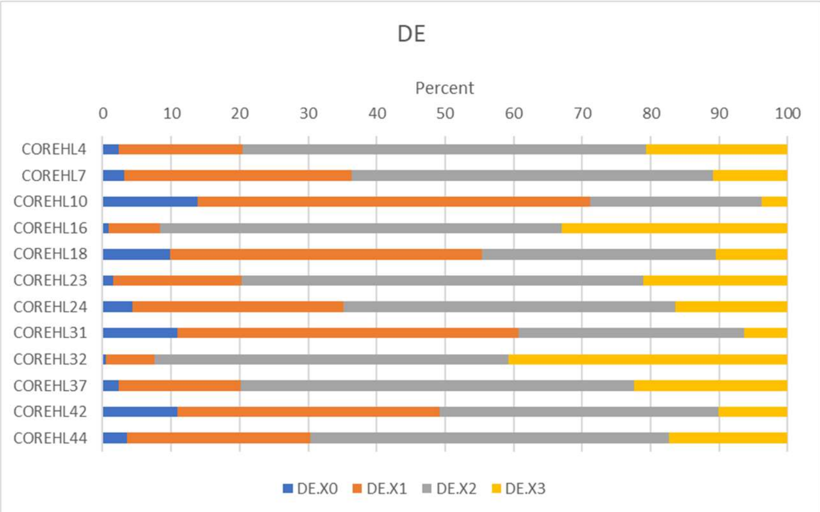
Figure A1: Frequency distributions for the response categories in the different countries

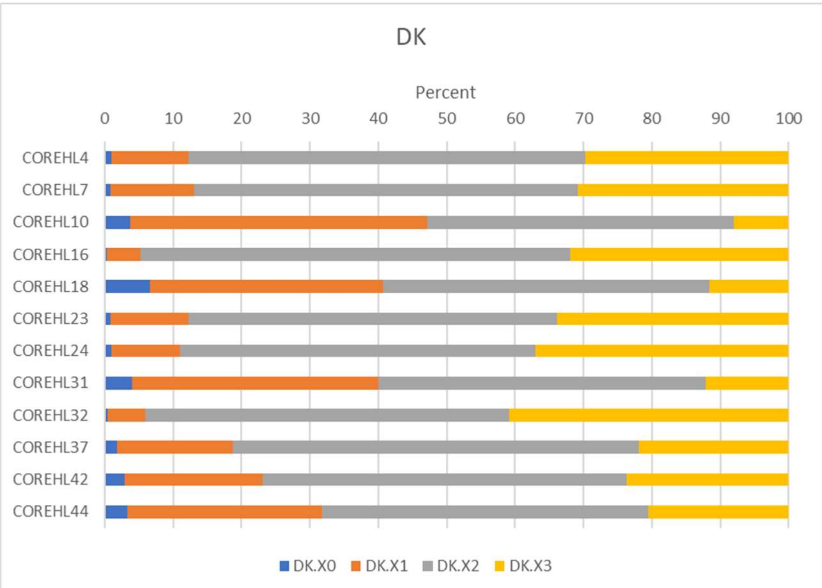


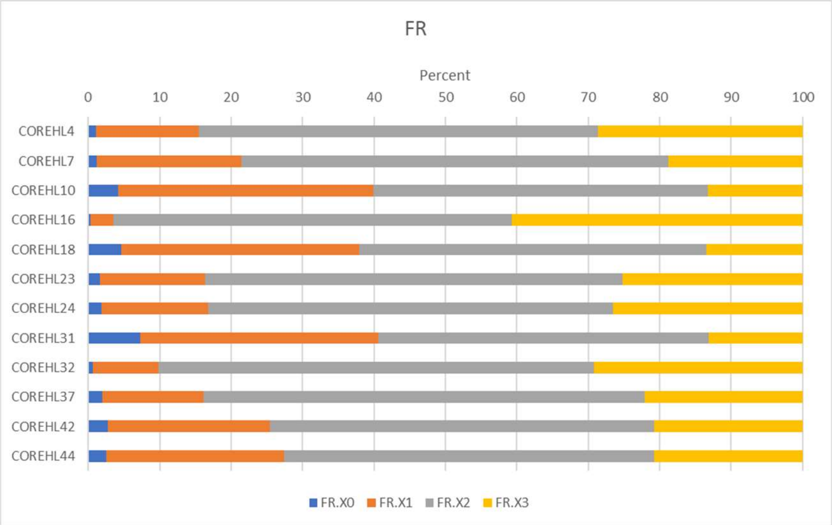


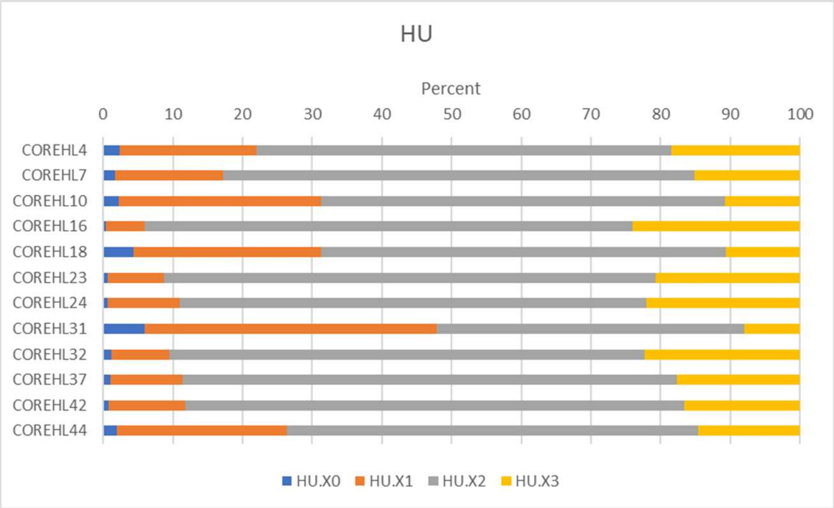


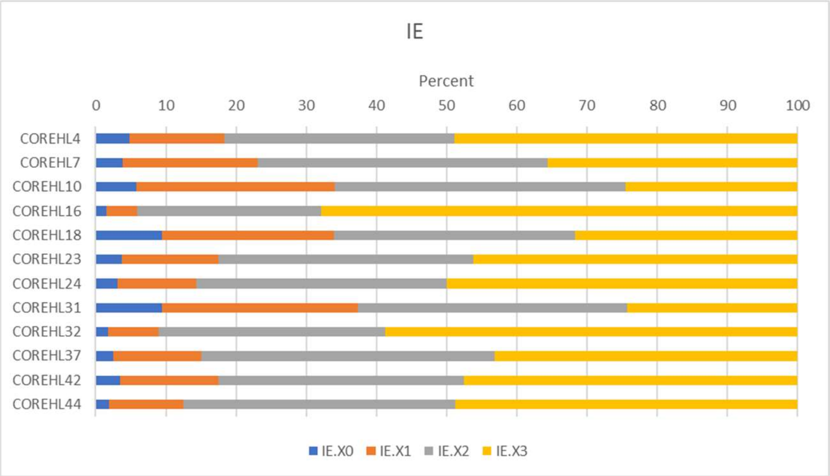

CZ
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
CZ.X0 CZ.X1 CZ.X2 CZ.X3

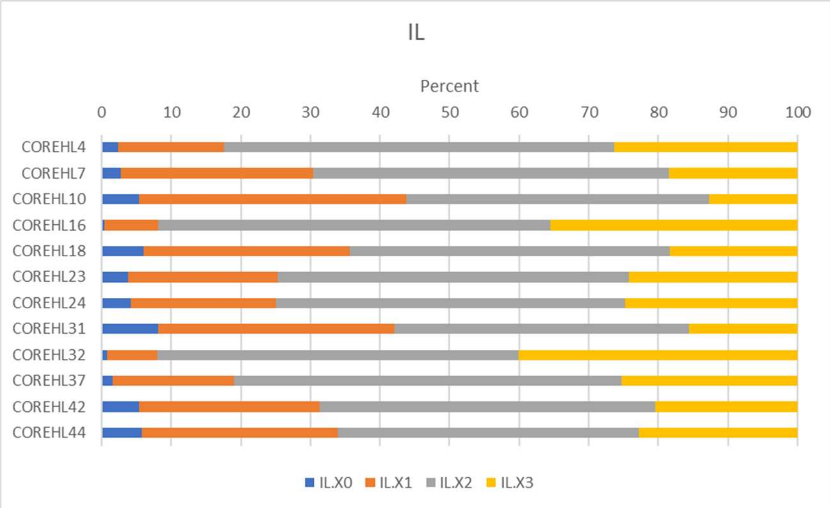

DE
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
DE.X0 DE.X1 DE.X2 DE.X3

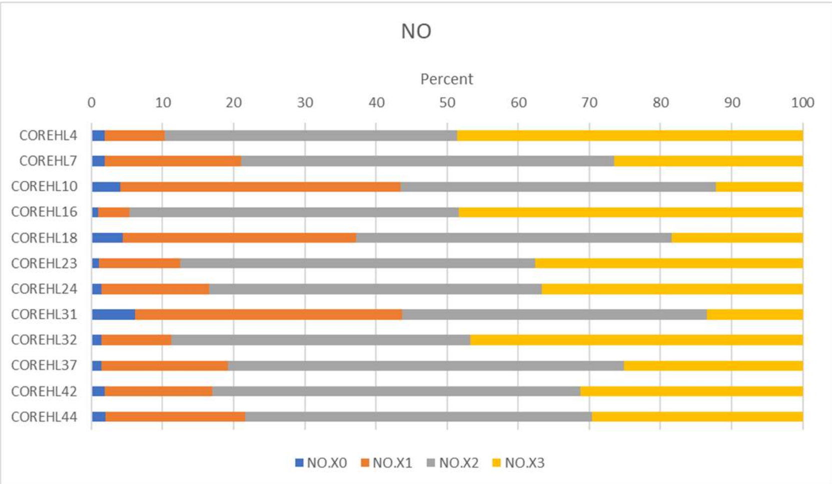

DK
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
DK.X0 DK.X1 DK.X2 DK.X3

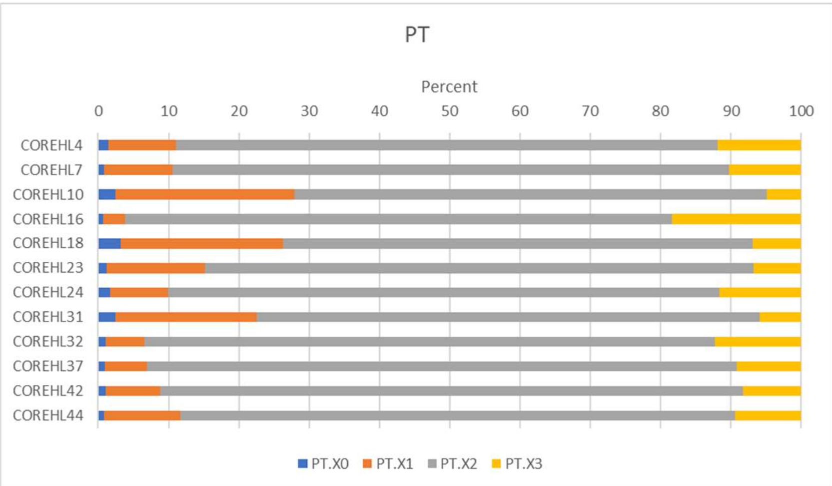

FR
Percent
0
10
20
30
40
50
60
70
80
90
100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
FR.X0
FR.X1
FR.X2
FR.X3

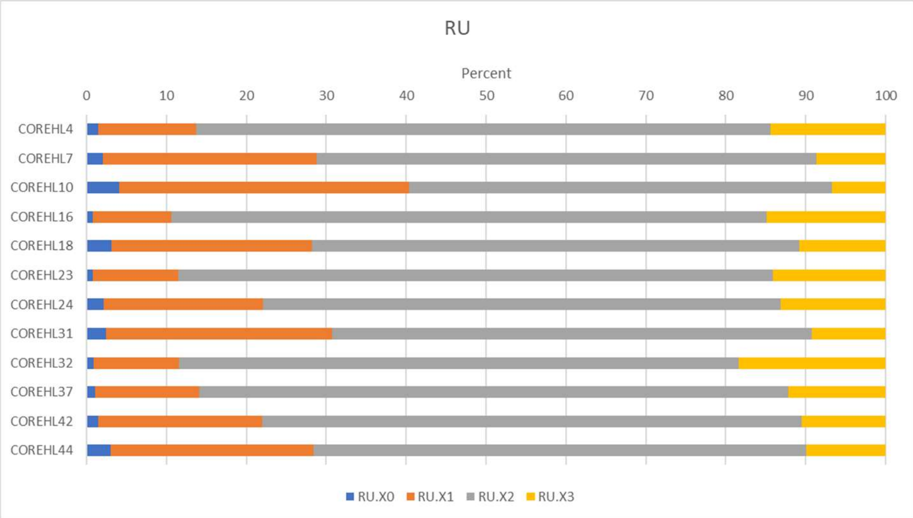

HU
Percent
0
10
20
30
40
50
60
70
80
90
100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
HU.X0
HU.X1
HU.X2
HU.X3

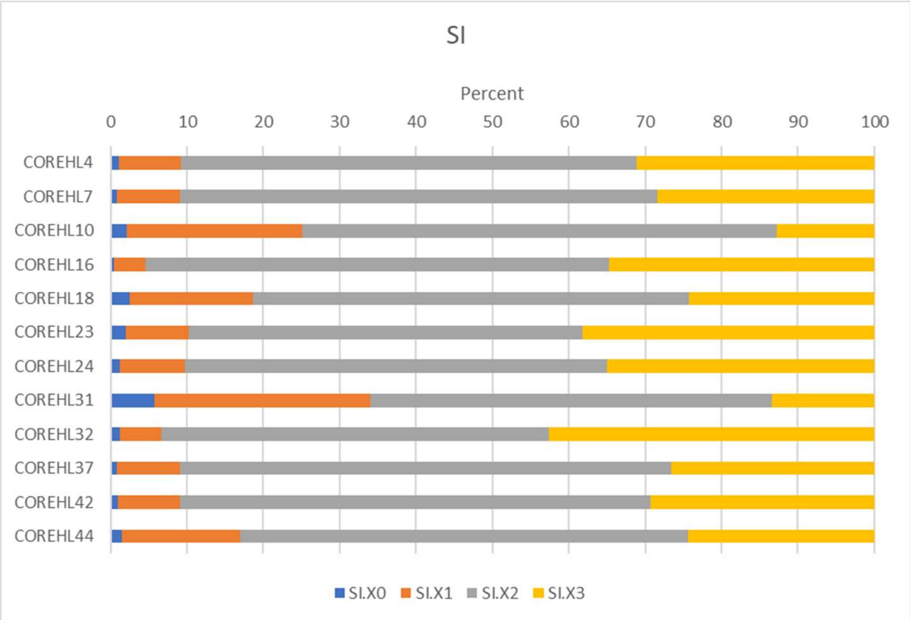

IE
Percent
0
10
20
30
40
50
60
70
80
90
100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
IE.X0
IE.X1
IE.X2
IE.X3

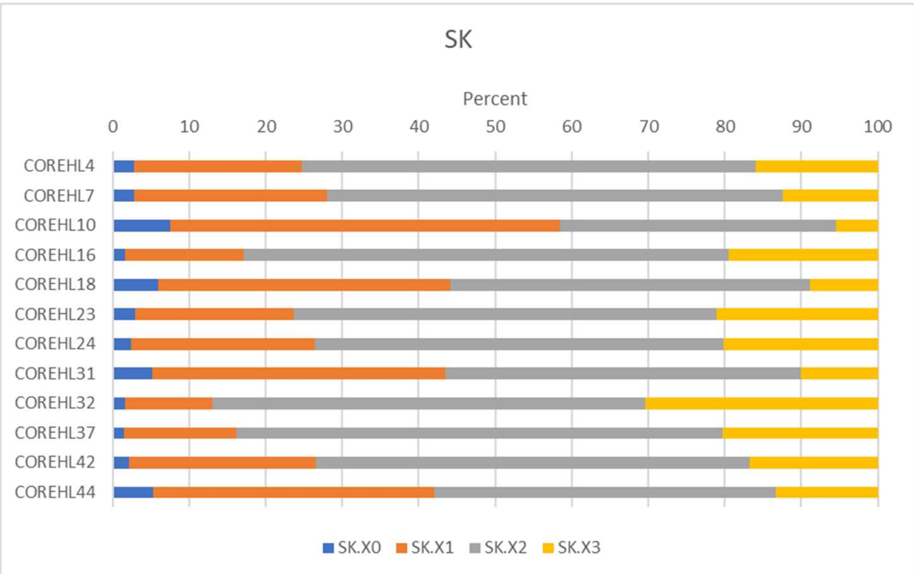

IL
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
IL.X0 IL.X1 IL.X2 IL.X3


NO
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
NO.X0 NO.X1 NO.X2 NO.X3


PT
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
PT.X0 PT.X1 PT.X2 PT.X3


RU
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
RU.X0 RU.X1 RU.X2 RU.X3


SI
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
SI.X0 SI.X1 SI.X2 SI.X3


SK
Percent
0 10 20 30 40 50 60 70 80 90 100
COREHL4
COREHL7
COREHL10
COREHL16
COREHL18
COREHL23
COREHL24
COREHL31
COREHL32
COREHL37
COREHL42
COREHL44
SK.X0 SK.X1 SK.X2 SK.X3