# Rasch analyses of data collected in 17 countries

## - A technical report to support decision-making within the M-POHL consortium

International Population Health Literacy Survey 2019-2021 ($HLS_{19}$)

13 October 2021



Øystein Guttersrud, Christopher Le, Kjell Sverre Pettersen, Hanne Søberg Finbråten

## Content

# INTRODUCTION

The National Study Team (NST) of $HLS_{19}$ for Norway offered to support the analyses of $HLS_{19}$ data by providing statistical analyses of assessment scales to test data-model fit by using Item-Response Theory procedures and Rasch modelling. **The Norwegian NST provides analyses of $HLS_{19}$-Q12 and the optional scales measuring navigation, communication and digital health literacy**. We have also added analyses of the 4-item **vaccination literacy** measure. As few countries applied $HLS_{19}$-Q47, and the international report will emphasize $HLS_{19}$-Q12, analyses of $HLS_{19}$-Q16 and the domain-specific subscales for HC, HP and DP are not reported.

# BACKGROUND

In this background section, we shortly introduce concepts, ideas and statistical models applied in the Analyses chapter.

## *Tests of fit*

For a dichotomously scored multiple choice item, a basic test of fit is the difference between the observed proportion 'correct' and the Rasch dichotomous model (Rasch, 1960) predicted proportion or residual. Based on test score sums, we group test takers into *G* number of groups or 'class intervals'.

A **z-fit residual** is the standardised difference between the observed and model predicted or expected number 'correct' in the class intervals (z-fit < -3.0 may indicate an over-discriminating item, and z-fit > 3.0 may indicate an under-discriminating item). To get a fit index of an item as a whole, these residuals are squared and added up. In the software package RUMM2030plus (Andrich & Sheridan, 2019) this gives an approximate **chi-square** distribution on $df = G\text{-}1$ degrees of freedom (and the overall chi-square based test of fit for $J$ items has $df = J \times (G\text{-}1)$. The 'chi-square **probability'** reports the 'probability of observing the estimated chi-square value' given good data-model fit ($p > .05$). Repeating the significance tests increase the probability of observing significant values or 'misfit items', and to counteract this effect we applied Bonferroni adjusted chi-square probabilities: $.05/J = .004$ for $J = 12$ items (Bland & Altman, 1995).

The **infit** fit index is an information- weighted or variance-weighted fit residual with expected value equal to 1 (infit < 1 indicates an over-discriminating or over-fitting item, and infit > 1 indicates an under-discriminating item or under-fitting item). For high stakes tests, such as exams, .8 > infit < 1.2. For $HLS_{19}$ measuring at the population level we may view .7 > infit < 1.3 as sufficient. Infit is 'inlier sensitive', and we may say it put more emphasis on 'targeted observations' and pay less attention to 'extreme observations'. Outfit, which we not reported, is 'outlier sensitive' and pay more attention to 'extreme observations'

Each $HLS_{19}$ measurement scale consists of a set of 'rating scale items' with **four** ordered response categories ('very difficult' – 'very easy'). Normally, we score these items 0–3 points, and we expect the scores to be ordered. We may 'generalize' the idea of fit indexes to these ordered polytomous items.

## *Sample size for the item calibration and data-model fit stage*

In general, the greater the sample size, the more powerful the test of fit that the responses do not fit the model. The largest sample size in $HLS_{19}$ is more than 5500 persons, and these data will most likely not fit any Rasch-type model.

With $k = 4$ response categories there are $k\text{-}1 = 3$ thresholds to be estimated. A rule of thumb based on 'substantial experience and simulation' is that the sample for data-model fit analyses should increase

as the number of item thresholds increase, and a reasonable ratio is **between 20 and 30 persons** for each threshold (Andrich, 2011, p. 7). For example, for a scale consisting of *J* = 12 items using a 4-point rating scale, we may therefore reduce the sample size / select a 'random sample' between 720 respondents (12 items x 3 thresholds x 20 persons = 720 respondents) and 1080 respondents (12 x 3 x 30 = 1080). We did this for each $HLS_{19}$ participating country. However, a smaller sample size may give meaningful results and identify anomalous items.

## *Unidimensional polytomous Rasch models and IRT models*

Masters (1982) formulated the polytomous Rasch partial credit model (**PCM**) by using the Rasch dichotomous model, and we may view the dichotomous Rasch model as a special case of PCM in which the number of categories is two. The rating scale parameterisation of the PCM (RSM; Andrich, 1978) constrains the Rasch-Andrich thresholds or 'step difficulties' to be equidistant across the items. We may consider using the RSM when item responses are elicited by a common set of behavioural anchors, such as 'strongly disagree' – 'strongly agree' or 'very difficult' – 'very easy' as in $HLS_{19}$-Q12, but the PCM is more flexible.

Following Masters, Muraki (1992) constructed the generalized PCM (**GPCM**) by using the Birnbaum dichotomous two-parameter logistic model (Birnbaum, 1968). Mathematically, PCM therefore follows from GPCM by fixing the item discrimination parameters $a_i$ to 1 (or $A_i = a_i D = 1$, where D equals 1.0 (pure logistic model) or 1.7 (logistic approximation)). We may say that GPCM relaxes the Rasch assumption of equal discrimination across items.

PCM and GPCM are 'divide-by-total-models' (Thissen & Steinberg, 1986) or 'adjacent categories models' estimating Rasch-Andrich thresholds. We did not consider any 'difference models' (Thissen & Steinberg, 1986) with a Thurstonian cumulative boundary measurement approach, such as the Samejima model (Samejima, 1969).

## *Why we prefer Rasch type models*

Specific objectivity – the separability of the item location $b_i$ and person location $\theta_j$ parameters, which implies sufficiency – the existence of the minimal sufficient statistics of the response data matrix, are distinct mathematical properties of the family of Rasch models (Wright & Stone, 1979, p. 20). When data fit Rasch models, we can defend summing up the raw scores. IRT models, which use the data matrix to estimate item discrimination parameters $a_i$ do not share these properties, and we would need to use weighted raw score (weighted with the discrimination parameter). However, this procedure does not meet the sufficiency assumption. As opposed to IRT models, Rasch models do not allow item characteristic curves to intersect and therefore meet the requirement of invariance.

The features of Rasch models permit a specialized parameter estimation procedure – conditional maximum likelihood estimation, which offers unbiased *item* estimates (Muraki, 1992, p. 160). Weighted maximum likelihood produces unbiased *person* estimates (Warm, 1989).

## *Model comparison*

The difference in deviance or -2*LL* from two hierarchically, nested models is distributed as a chi-square with *df* equal to the difference in the number of *df* between the 'full' and the 'reduced' model (de Ayala, 2009, p. 140). We may therefore compare the more complex or less constrained GPCM to the less complex or more constrained PCM by using likelihood ratio test (LRT), where PCM is 'nested within' GPCM. To compare GPCM to PCM by using LRT we must estimate the models on the identical sample. We may also calculate the relative reduction in deviance from PCM to GPCM (de Ayala, 2009, p. 141), quite similar to comparing R squared for nested regression models.

If data do not sufficiently fit the Rasch PCM, we may choose the GPCM. However, then a respondent's score sum is not the simple raw score but the sum of weighted scores.

### *Software*

The PCM was fit using the software RUMM2030plus (Andrich & Sheridan, 2019) and the software Conquest5 (Adams et al., 2020). We fitted and compared the GPCM to PCM by using the software Xcalibre 4.2.2 (Guyer & Thompson, 2014).

### *Mode – method of data collection*

In $HLS_{19}$, participating countries selected between different methods of data collection or modes

- computer-assisted telephone interviewing (CATI) – a telephone surveying technique in which the interviewer follows a script provided by a software application
- computer-assisted web interviewing (CAWI) – an internet surveying technique in which the 'interviewee' follows a script provided in a website
- computer-assisted personal interviewing (CAPI) – a personal interview technique in which the respondent or interviewer uses an electronic device to fill the answers into the questionnaire
- paper-assisted personal interviewing (PAPI) or 'paper and pencil interviewing' – a personal interview where the pollster has a printed-out questionnaire, reads the question to the respondent and fills the answers into the questionnaire

# ANALYSES

In this section, we emphasize the $HLS_{19}$-Q12 but all ideas and comments apply to all scales (digital, communication and navigation health literacy). If not explicitly stated otherwise, all analyses refer to the PCM with raw data scored 0–3 points reflecting the 4-point rating scale 'very difficult' – 'very easy'.

### *Using data for quality assurance*

<u>*Overall data-model fit*</u>: Testing data up against the PCM for country wise samples with 20 persons per threshold, Table A1 displays good overall data-model fit for $HLS_{19}$-Q12 in Austria (CATI), Denmark, Germany, Israel (CAWI), Norway, Slovakia and Switzerland. This conclusion is based on $\chi^2(df = 84, n = 720)$, $\boldsymbol{p > .05}$. Table A1 displays sufficient overall data-model fit for $HLS_{19}$-Q12 in Austria (CAWI), Belgium, Czechia (CAWI and CATI) and Ireland. This conclusion is based on $\chi^2(df = 84, n = 720)$, $\boldsymbol{p > .01}$. Reducing sample size down to $n = 360$ or 10 persons per threshold, France, Hungary, Russia and Slovenia display acceptable overall data-model fit. Owing to the relatively large overall $\chi^2$, the Portuguese data display acceptable overall data-model fit for $n = 360$, that is, $\chi^2(df = 84, n = 360)$, $\boldsymbol{p > .01}$.

**[**Fitting the more complex model GPCM to the Portuguese data, results in a relative reduction in deviance *-2LL* of (12892-12458)/12892 = .034 (see *D* = -2*LL* in Table A1). The interpretation is that the GPCM results in an improvement of fit of 3.4% over the PCM (see the column RelRed or 'relative reduction' in Table A1). The difference in deviance between PCM and GPCM is 12892-12458 = 434 and change in deviance is asymptotically chi-square distributed with *df* equal to the difference in estimated parameters. Here *df* = 12 as the GPCM estimates one discrimination parameter for each of 12 items. As the PCM is nested within GPCM, we can test the size of the change in deviance by using likelihood ratio test: $\chi^2(df = 12) = 434$, $p << .01$. The test indicates that the data fit significantly better to the GPCM than the PCM. In Table A1 we have reported results for GPCM only for the five countries with poorest fit to PCM**]**. We put this section in square brackets to specify that we discuss the GPCM only here.

*Individual item data-model fit*: Good overall data-model fit usually implies good data-model fit for single items. Using the German data as example, we see from Table A2 that all $HLS_{19}$-Q12 items sufficiently fit the PCM. Item 37 'to understand advice concerning your health from family or friends' under-discriminates somewhat (z-fit > 3.0 and significant chi-square, but the inlier sensitive infit < 1.2). In Portugal, the $HLS_{19}$-Q12 items tend to over-discriminate (Table A2). If we leave the Rasch-paradigm and steps into the IRT-paradigm, we would interpret the Portuguese data as quite 'strong'. We saw that the IRT-model GPCM, which models item discrimination, improved data-model fit for the Portuguese data.

*Ordering of response categories*: Estimating Rasch-Andrich thresholds, we can evaluate whether thresholds are ordered. Ordering of thresholds are easiest understood by looking at the 'category probability curves' for $HLS_{19}$-Q12 item 16 in Figure A. The red curve termed '2' indicates the probability of ticking off in category 2 'difficult' as a function of respondents' standing on the latent trait 'health literacy' along the x-axis. The right diagram shows slightly reversed thresholds for the Irish data and indicates that category 2 not is the most likely for any health literacy level. $HLS_{19}$-Q12 item 16 displays unordered thresholds in the Belgian, Irish, and Norwegian data (not significant in the Norwegian data). For example, 'your pharmacist' does not apply to the Norwegian context. In the Austrian data, we observed *in*significant unordered thresholds for $HLS_{19}$-Q12 item 4. To sum up, only $HLS_{19}$-Q12 item 16 displayed unordered thresholds in as few as two countries – a remarkably good result. Therefore, there is **no need to rescore**, or more extremely, **dichotomize** the $HLS_{19}$-Q12 items. Dichotomising the items will lead to excessive loss of information, reduced variance / lower reliability and significantly less variance in health literacy to explain by regression models.

The idea behind the suggestion of **completely removing** the phrases anchored with the **two middle categories**, which ended up in removing the word '**fairly**', was to reduce the chance of observing unordered thresholds. To decide whether removing 'fairly' actually was an effective strategy, we need to re-analyse the HLS-EU data and compare the occurrences of unordered thresholds in the HLS-EU datasets to the $HLS_{19}$ datasets.

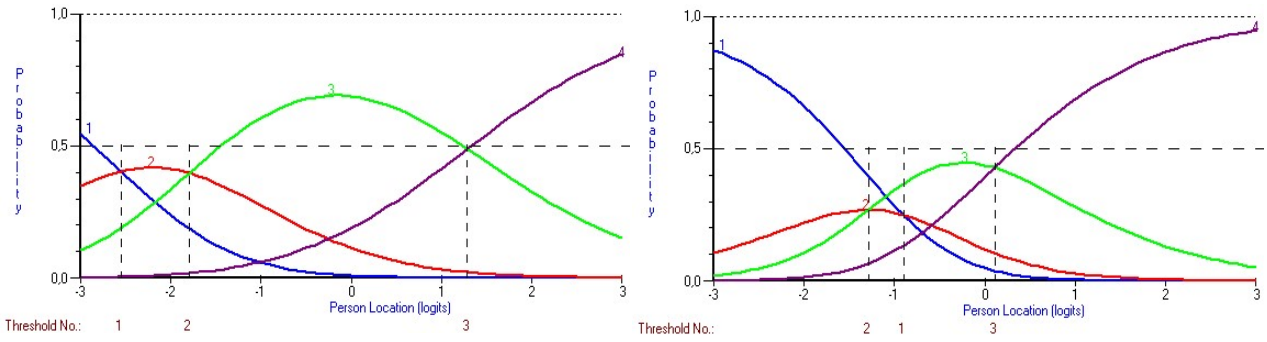


**Figure A.** Visualizing ordering of thresholds for $HLS_{19}$-Q12 item 16 'to act on advice from your doctor or pharmacist' for the German (left) and the Irish (right) data.

*Targeting and mean health literacy proficiency*: After calibrating the items, we estimated the person locations or person proficiency estimates by using Warms' weighted maximum likelihood estimation (Warm, 1989). Table A1 reports the mean Rasch-based person estimates in logits. The overall 'difficulty' of the $HLS_{19}$-Q12 items fit well to the overall health literacy level of the Belgian (mean = .62) and German (mean = .65) samples when using CAWI and PAPI, respectively. Figure B1 visualizes the concept of targeting. In, for example, Austria the $HLS_{19}$-Q12 scale is somewhat 'out of target'. We may compare measuring health literacy in Austria by using $HLS_{19}$-Q12 to a mathematics test that was too easy for students.

Figure B2 displays the distributions of the two comparable Austrian samples (CATI and CAWI) when the samples are merged to obtain a common 'point of reference' for the $HLS_{19}$-Q12 measurement scale. In line with literature (Christian et al., 2008; Lugtig et al., 2011; Ye et al., 2011), CATI results in significantly higher mean health literacy level than CAWI do. Czechia / Czech Republic, Israel, Italy and Switzerland also collected data using both CAWI (large sample) and CATI (smaller sample), but the CATI and CAWI samples are *not* directly comparable within countries. Compared to the respective CAWI sample, the Czech and Italian CATI samples had an overweight of females and older people. In Italy, the $HLS_{19}$-Q12 CATI data displayed somewhat poor overall data-model fit. Owing to small sample size, we did not estimate the Swiss CATI sample.

Germany changed from CAPI in HLS-EU to PAPI in $HLS_{19}$, but we have no available data to assess how this influenced the German results. Slovenia and Bulgaria collected data using both CAPI and CAWI, and the Slovenian and Bulgarian data displayed somewhat poor overall data-model fit. The Slovenian CAPI sample had an overweight of older people with low education. The Bulgarian CAPI and CAWI samples were somewhat small with few old people and overweight of females.

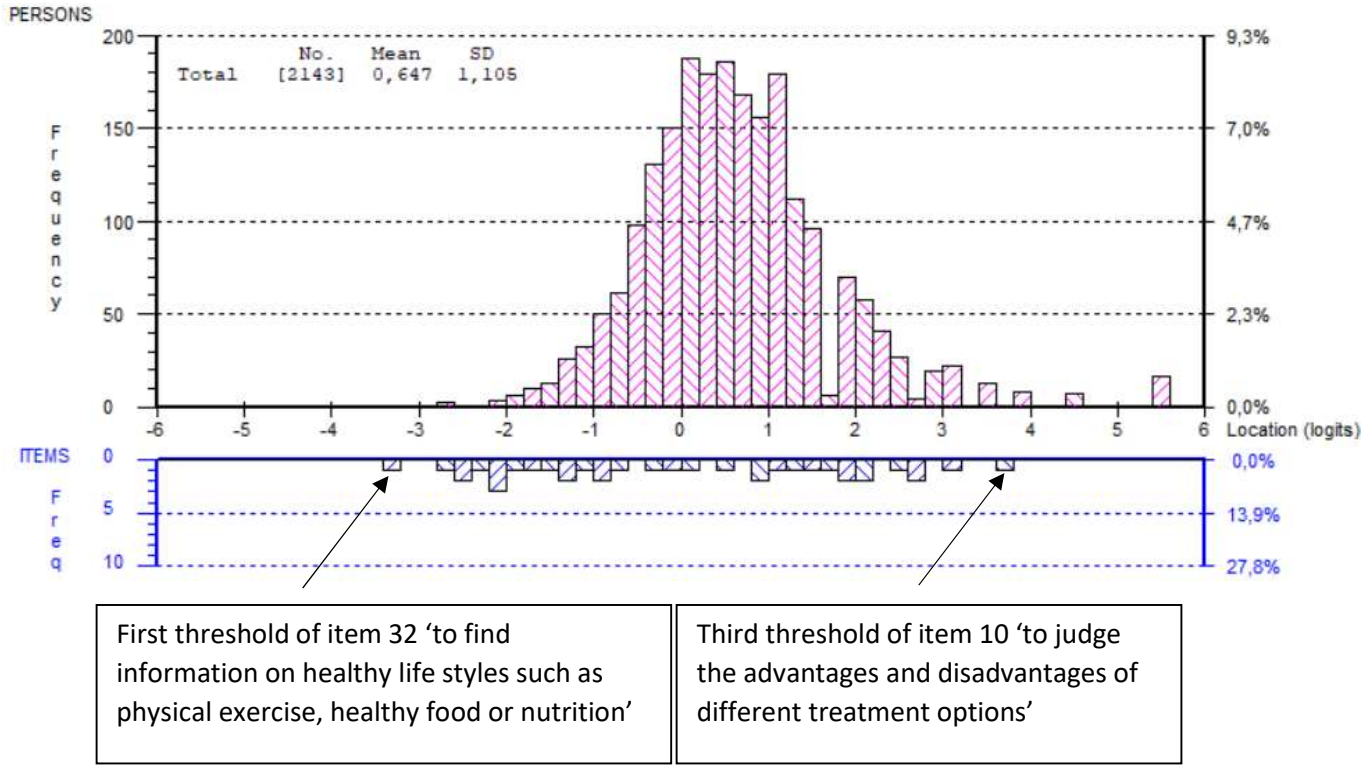


**Figure B1.** Visualizing the concept of targeting – how well the distribution of item thresholds (histogram below the x-axis) fits the distribution of person proficiency estimates (histogram above the x-axis) in the German $HLS_{19}$-Q12 data.

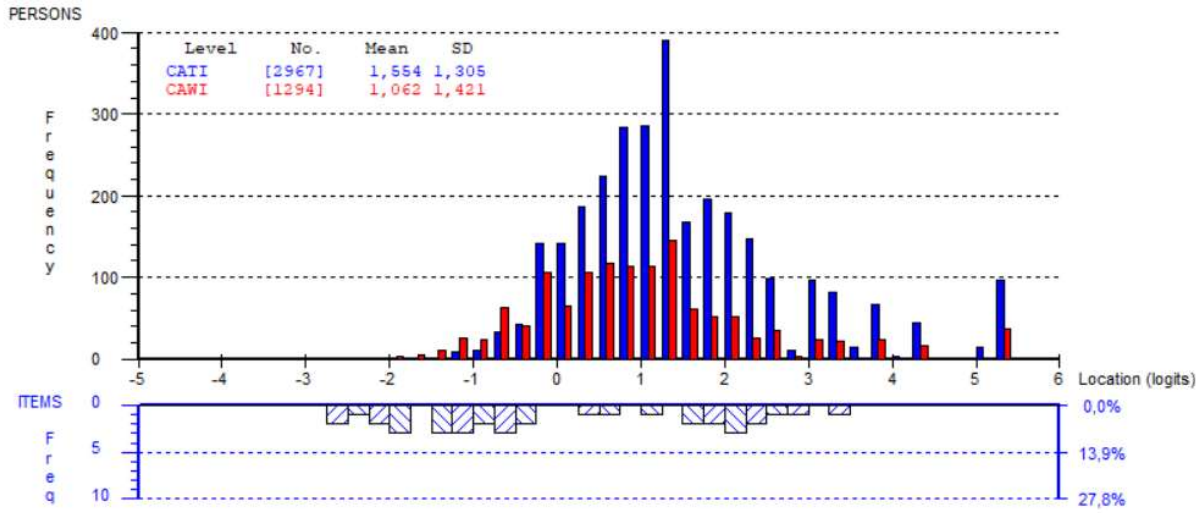


**Figure B2.** This figure is equivalent to Figure B1 but visualizes the distributions of person proficiency estimates (histograms above the x-axis) for CATI (blue) and CAWI (red) in the Austrian $HLS_{19}$-Q12 data, when the CATI and CAW data sets are merged. Austria collected two large and comparable samples. CATI results in better data-model fit (Table A1) and significantly higher scores on the outcome variable health literacy. The distribution of item thresholds (histogram below the x-axis) could have been better targeted to the distributions of person proficiency estimates (histograms above the x-axis).

*Dimensionality:* The $HLS_{19}$-Q12 items measure three health domains (health care, health promotion and disease prevention), and four cognitive domains (find, appraise, understand and apply). These different domains or aspects capture the complexity of the construct and increase the validity of the $HLS_{19}$-Q12 scale, but they inevitably bring multidimensionality into the measure.

There are several approaches to testing for unidimensionality. Using the software Conquest, we can estimate and, by likelihood ratio test, compare a multidimensional to the nested unidimensional model. We may also form subtests and observe drops in inflated reliability indices. A quick check is to form two subsets of items and estimate the proportion of respondents with significantly different person estimates on the two subsets. If the portion of significant dependent *t*-tests > .05 (theoretical subscales) or > .10 (empirical subscales based on principal component analysis (PCA) of Rasch residuals), the scale is not strictly unidimensional. Using the latter method with empirical subscales, we found no systematic subsets of items across countries. Table A1 shows that the percentage significant *t*-tests is below 10 % for each country (the column Dim (%)), varying between approximately 10 % in Denmark and 5.5 % in Norway. Therefore, $HLS_{19}$-Q12 seems sufficiently unidimensional.

*Reliability*: Scoring the 12 items using 4-point rating scale 0–3 points, each respondent has a sum score 0–36 points. We therefore expect a rather broad spread or separation of persons and, consequently, high reliability indexes (Person Separation Index (**PSI**) and **Cronbach's α**). We estimated PSI on original datasets with missing values, while Cronbach's α was estimated using only respondents with complete $HLS_{19}$-Q12 data. As Belgium submitted a complete dataset, PSI and Cronbach's α are very similar. Table A1 display sufficiently high reliability for all countries, on the assumption of unidimensional data.

*Response dependency*: If all 12 items were identical, the respondents would be stratified into four groups (or three groups if no one used the 'very difficult' category). These groups of respondents would be strongly separated based on significantly different score sums, and reliability indices would be close to 1. We check for 'too similar' or dependent items by estimating correlations between item Rasch model residuals. We used residual correlation > .3 as a target value for response dependency between items. Not reported here, but we found no evidence of response dependency or 'too similar' items

within the $HLS_{19}$-Q12scale. Another way to put it is that no pair of items shared variance over and beyond the latent trait 'health literacy'.

*Differential item functioning (DIF)*: To examine for DIF, we used two-way analysis of variance of standardised residuals (Andrich & Marais, 2019, p. 201). For DIF analyses, we categorised person factors as shown in Table 1.

**Table 1.** Person factor levels used for analysis of differential item functioning (DIF).

| Person factor | Levels (categories) | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| gender (CDET1) | male | female | |
| age (CDET2) | | | |
| - dichotomised | 18 to 45 years | 46 years and older | |
| - agecat1 | 18 to 25 years | 26 to 65 years | 66 years and older |
| - agecat2 | 18 to 45 years | 46 to 75 years | 76 years and older |
| education (CDET6) | ISCED 0–3 | ISCED 4–8 | |
| employment (CDET7) | employed | unemployed or retired | |
| pay bills (CDET11) | easy | difficult | |
| social level (CDET12) | level 1–4 | level 5–10 | |
| general health (CHSTAT1) | good or fair | bad | |

Several items displayed DIF when sample size was reduced to 1080 (see Table A2 in appendix), and for some items DIF was still evident when reducing the sample size to 720. One example is item 23 'understand information about recommended health screenings or examinations', which displayed DIF for 'employment' in a few countries (BE, FR, SI (CAWI)) and for respondent age in several countries (BE, CH, DK, FR, SI (CAWI and CAPI)). Using the Danish data, Figure C visualizes how item 23 displays DIF for respondent age (variable CDET2). The interpretation is that, despite same level of health literacy, older respondents tend to respond more often '(very) easy' on item 23 than younger respondents do. Young people may be less familiar with health screenings and have poorer understanding of information about these examinations. Unemployed or retired people, who possibly have health problems and therefore may be more exposed to this type of information, more often respond '(very) easy' on item 23 than employed respondents.

We observed several items displaying significant DIF for respondent age (items [country]: 32 [PT], 37 [CH],) 42 [DK]) and employment status (31 [BE], 32 [HU], 42 [DK], 37 [CH]) even when sample size was adjusted to 720. In addition, in the Belgian data, item 37 'to understand advice concerning your health from family or friends' displayed DIF for the factors 'pay bills' and 'health status'. In the Portuguese data item 10 'to judge the advantages and disadvantages of different treatment options 'displayed DIF for 'social levels' and in the Slovenian data item 31 'to decide how you can protect yourself from illness using information from the mass media' displayed DIF for 'educational background'. Conditional on sample size 1080, we did not observe items displaying significant DIF in the Austrian, Russian or Slovakian samples.

Figure D shows an example on an item displaying DIF for country of residence, and this is a significant challenge in the $HLS_{19}$ data. **Further analyses will indicate whether we may compare results across smaller groups of countries, like German-speaking countries**.

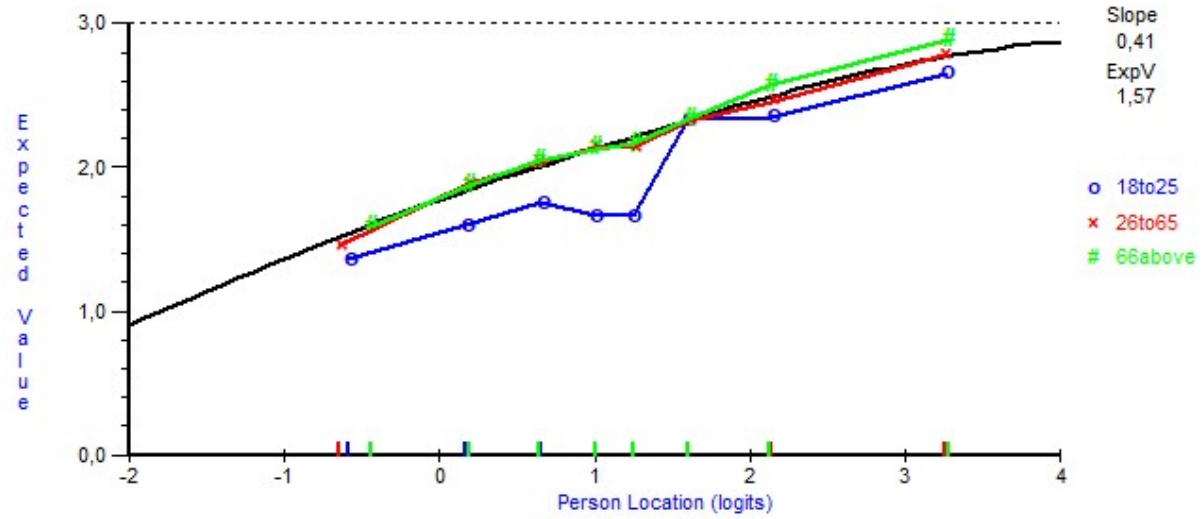


**Figure C.** Visualizing the concept of DIF – how people in Denmark with the same health literacy proficiency have a different probability of giving a certain response to $HLS_{19}$-Q12 item 23 depending on their age (person factor age (CDET2) with levels 18–25 years, 26–65 years and 66 years and older).

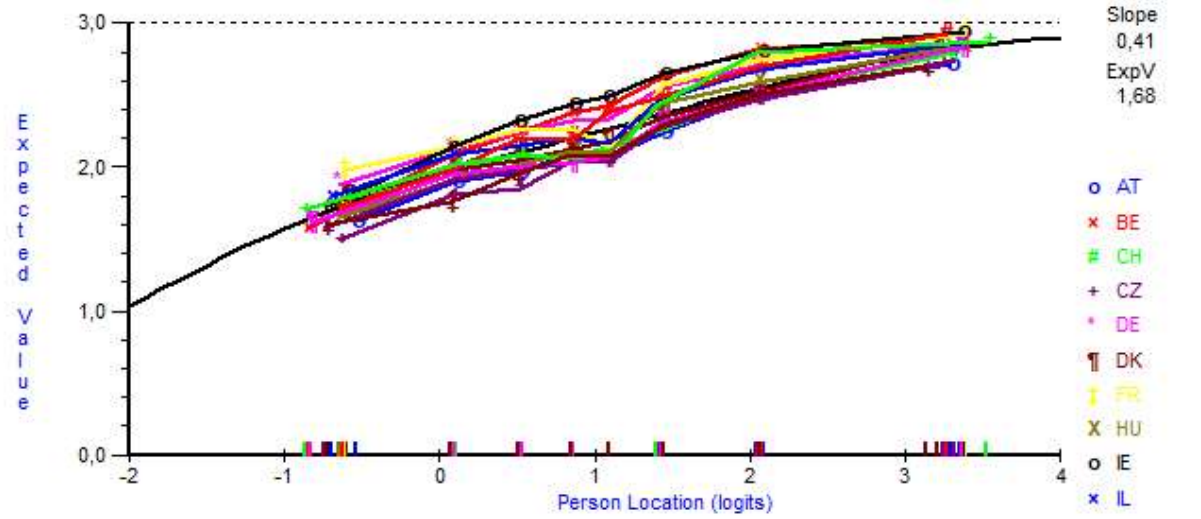


**Figure D.** Visualizing the concept of DIF – how people with the same health literacy proficiency have a different probability of giving a certain response to $HLS_{19}$-Q12 item 16 depending on country of residence.

*Conclusion*: To conclude, despite some deviancy the overall picture is that **the country wise $HLS_{19}$-Q12 data seem to have acceptable quality**. Having said that, the analysis provided is an excellent source for initiatives to revise the $HLS_{19}$-Q12 scale. Items displaying DIF between countries are a challenge for comparative analyses. We specifically mention that the theoretical subscales of the navigation health literacy scale seem to bring some multidimensionality into the measure.

## *Estimating individual health literacy proficiency estimates*

Using Rasch modelling with FIML (full information maximum likelihood estimation), which is not an imputation method; we can easily estimate a score for a respondent with missing data. Owing to missing data or 'lack of information', respondents with missing data may have larger standard error of estimate. The same is true for respondents with very high or low health literacy scores, as there are few 'hard and easy' items providing information at low and high locations.

The software output in Figure B displays that we estimated a Rasch-based health literacy score for each of 2143 German respondents. Therefore, there is no need to provide a '**procedure for constructing scores**' as Rasch modelling with FIML elegantly solves the challenge. Subsequently, there is no need to '**change the procedure for constructing the scores'** as a '**consequence of the change of wording of the response categories**'. Above we indicated that there is no need to rescore – at least not dichotomize – any $HLS_{19}$-Q12 items as the response categories worked very well.

As opposed to any form of raw score, the Rasch-based $HLS_{19}$-Q12 health literacy scores have 'interval scale property' and an underlying continuous latent trait. These estimates meet the regression assumption of a continuous dependent variable; and the interval property makes the regression parameter estimates valid for all health literacy levels. The standard errors associated with the estimates are conceptualized as a component of the regression residual $e_i$, where Y = Y_hat + $e_i$. If we use Rasch-based estimates for independent variables in regression, the error of measurement tends to attenuate the parameter estimate. Structural equation modelling solves this problem.

## *Estimating progression/ change over time by using anchors or item linking*

We agree that it is difficult to demonstrate how the **reworded questions** (items revised between HLS-EU and $HLS_{19}$) have affected the data, but that is not a relevant discussion. There are different approaches to estimating change in overall health literacy from HLS-EU to $HLS_{19}$, but we must treat the revised items as 'new' items.

*Anchoring*: One approach is to use the identical items (items not revised between HLS-EU and $HLS_{19}$) as 'anchors'. In the anchored item analysis, we import item parameters (the item principal components) for some or all items 'in a scale' from a previous analysis (e.g., the HLS-EU study) to estimate person abilities for the $HLS_{19}$ cohort. Then, we estimate person abilities for the $HLS_{19}$ cohort in a prior frame of reference (the HLS-EU cohort).

*Linking*: Another approach is to merge the country wise HLS-EU and $HLS_{19}$ datasets, link the few identical items, and set HLS-EU specific items to 'systematic missing' for the $HLS_{19}$ cohort and vice versa. Then, the data set involves blocks of systematically missing which are the result of the test design rather than respondents' 'missing' responses. Using FIML, like pairwise maximum likelihood algorithm, the calculation of the sufficient statistics for item parameters allows for missing data.

Then, we can estimate the mean health literacy proficiency for the EU-HLS respondents and the mean health literacy proficiency for the $HLS_{19}$ respondents based on a common origin or 'point of zero'. By comparing these estimates, we have measured change. The latter method is valid as the EU-HLS respondents and $HLS_{19}$ respondents are different/independent individuals (not a pre and post design).

If any countries applied HLS-EU-Q16 and later applied the $HLS_{19}$-Q12, they may estimate change by linking these two scales by item q4 – the one common item that we not revised between HLS-EU and $HLS_{19}$.

A relevant discussion is change of 'mode' from HLS-EU to $HLS_{19}$. Eight countries took part in the 2011 HLS-EU study (Austria, Bulgaria, Germany, Greece, Ireland, the Netherlands, Poland and Spain), and of these only three countries (Austria, Germany and Ireland) participated in the $HLS_{19}$ study. As these three countries changed mode or data collection method from the HLS-EU to $HLS_{19}$ (Austria changed from CAPI to CATI, Germany from CAPI to PAPI, Ireland from PAPI to CAWI) measuring change is practically impossible. Further, some items were revised or reworded, and the word 'fairly' was removed from the two central response categories. Possible effects of all these changes (mode, rewording and changing the central response categories) cannot be isolated in the available data.

However, other countries may have repeatedly collected data over time and can measure change. For example, Norway collected data using CATI for both the HLS-EU-Q47 and the $HLS_{19}$-Q47 questionnaires. Based on field trials of the HLS-EU-Q47, where several items displayed reversed thresholds, Norway modified the middle response categories in both questionnaires (HLS-EU-Q47 and $HLS_{19}$-Q47). Hence, these data sets may be linked using identical items (items not revised between HLS-EU and $HLS_{19}$).

## *Using $HLS_{19}$ data to improve future assessments*

*At country level:* For Norway, Table A1 displays good overall fit for $HLS_{19}$-Q12, and Table A2 in the appendix displays reasonably good fit at the single item level. However, item 23 'to understand information about recommended health screenings or examinations' discriminates quite strongly between respondents with low versus high health literacy (large negative z-fit residual, large chi-square value and infit below .90). Over-fitting items are, of course, more a theoretical than an empirical problem. In this case, we may explain the deviation by a national adaption of item 23. Based on this information, the Norwegian group should discuss whether their national adaptation of item 23 could have caused the strong discrimination. As mentioned, item 23 displays DIF in several countries. DIF may be a significant challenge, especially in regression models where we use background variables to explain variance in health literacy. **All $HLS_{19}$ participating countries may use the single item information provided in the appendix as a basis for item revision**.

*At group level:* A greater concern in the Norwegian data is item 31 'to decide how you can protect yourself from illness using information from the mass media', which discriminates somewhat poorly and tend to under-fit the PCM (large positive z-fit residual, large chi-square and infit approaching 1.2). Item 31 discriminates somewhat poorly also in Austria, Belgium, Switzerland, Hungary, Ireland and Slovenia – half of the participating countries. Thus, it seems to be a systematic problem with item 31 across countries. M-POHL may therefore discuss whether item 30 should replace item 31 in the Q12 short version. Item 30 is included in the original Q12 short version (Finbråten et al., 2018).

**M-POHL should discuss** why certain items display poor measurement properties across countries/health systems and use this information in item revision. **M-POHL should also discuss** item selection – whether we may benefit from replacing specific $HLS_{19}$-Q12 items with suitable $HLS_{19}$-Q47 items.

## *Using $HLS_{19}$ data to inform health policy*

We have evidence that different modes, as CATI and CAWI, influence the mean health literacy measured by $HLS_{19}$-Q12 (see above). It follows that **we cannot define one cut-off value for 'inadequate' health literacy that is valid across different modes**. For example, in Israel and Czechia the estimated proportion with 'inadequate' health literacy in the CATI data may be different from /lower than in the CAWI data. Also, whether a certain health literacy level is 'sufficient' depend on contextual factors, like the functioning and structure of the health service in a specific country. Further, different countries have different political aims and goals, and the $HLS_{19}$ results should align with and inform that policy. Then, policy makers will view $HLS_{19}$ as relevant and central to health policy development in their 'local' or 'social' context.

A simple and straightforward method to estimate cut-off values for 'significantly different health literacy scores' for a specific country, is to use the Rasch-based estimates with pooled standard error. For example, if the sum score 20 points is associated with the proficiency estimate -2.33 logits (SE = 0.49) and the sum score 27 points is associated with the proficiency estimate -0.75 logits (SE = 0.48), we can conclude that -0.75 is outside the confidence interval -2.33+(2*SQRT(0.49^2+0.48^2)) = -0.96. Therefore, the sum score 27 points is significantly different from the sum score 20 points (while the

sum score 26 is not). Repeating the procedure, we can estimate a set of cut-off values, and we can estimate the percentage of respondents at each 'level'.

When the uncentralized item thresholds are ordered, we can ascribe the threshold associated with the response 'easy' for each item to specific points along the latent trait. Groupings of item content at and slightly above each cut-off point define bands or levels of achievement, and these levels of achievement may refer to knowledge and skills necessary to realise specific local health policy goals. We may use the ordered uncentralized thresholds to build up an achievement scale as their order and location on the scale reflect their increasing 'difficulty' (e.g. Van Wyke & Andrich, 2006).

In Norway, we applied this method. A main result was that knowledge and skills necessary to realise the political aim referred to as 'the patient's health service' was associated with health literacy at level 2. As 33 % scored below level 2, we concluded that one third of the population may lack the knowledge and skills necessary to meet the expectations of and realise this central political aim in Norway. Hence, the $HLS_{19}$ survey explicitly informed Norwegian health policy.

Table B1, Table C1a, Table C1b and Table D1 provide similar information as Table A1 but for other $HLS_{19}$ scales. Table A2, Table B2, Table C2a, Table C2b and Table D2 in the appendix provide single item statistics for the different $HLS_{19}$ scales.

**Table A1.** $HLS_{19}$-Q12 overall analyses. If not stated otherwise, all analyses refer to the PCM

| Country | $\chi^2$, *p* | Mode | Mean[g] | Reliability | | Dim (%) | D = -2*LL* | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | *α* | *PSI* | | PCM | GPCM | RelRed (%) |
| **Austria** | 72.1, .82 | CATI | 1.55[h] | .84 | .82 | 7.5 | | | |
| **Austria**[a] | 115.5, .01* | CAWI | 1.06[h] | .86 | .85 | 10.1 | | | |
| **Belgium** | 109.4, .03* | CAWI | .62 | .88 | .88 | 7.7 | | | |
| **Bulgaria**[e2] | 141.2, .00** | **CAPI** | .70 | .81 | .79 | 6.7 | | | |
| **Bulgaria**[e2] | 238.2, .00** | **CAWI** | .90 | .85 | .83 | 11.0 | | | |
| **Czechia**[b] | 110.3, .03* | CATI | 1.12[h] | .82 | .79 | 7.0 | | | |
| **Czechia** | 112.8, .02* | CAWI | .83[h] | .84 | .84 | 6.4 | | | |
| **Denmark** | 79.9, .61 | CAWI | 1.38 | .86 | .85 | 10.0 | | | |
| **France**[f] | 176.1, .00** | CAWI | 1.32 | .89 | .88 | 6.5 | 35283 | 34765 | 1.5 |
| **Germany** | 76.4, .71 | PAPI | .65 | .80 | .81 | 8.7 | | | |
| **Hungary**[f] | 137.0, .00** | CATI | 1.21 | .84 | .83 | 9.3 | 20657 | 20524 | .6 |
| **Ireland** | 113.5, .02* | CAWI | 1.22 | .82 | .79 | 5.9 | | | |
| **Israel**[c] | 111.2, .03* | CATI | 1.42[h] | | .85 | | | | |
| **Israel** | 98.5, .13 | CAWI | .95[h] | .87 | .87 | 8.1 | | | |
| **Italy** | 159.7, .00** | CATI | .71 | .84 | .81 | 9.1 | | | |
| **Italy** | 66.2, .92 | CAWI | .81 | .90 | .88 | 8.0 | | | |
| **Norway** | 91.5, .27 | CATI | 1.29 | .84 | .83 | 5.5 | | | |
| **Portugal**[f] | 225.8, .00** | CATI | 1.26 | .90 | .82 | 5.8 | 12892 | 12458 | 3.4 |
| **Russia**[f] | 135.3, .00** | PAPI | 1.11 | .90 | .87 | 7.0 | 81446 | 79763 | 2.1 |
| **Slovakia** | 81.2, .56 | CAPI | .88 | .88 | .88 | 9.1 | | | |
| **Slovenia**[f] | 145.2, .00** | **CAPI** | 1,67 | .91 | .88 | 7.3 | | | |
| **Slovenia**[f] | 201.6, .00** | **CAWI** | 1.85 | .86 | .84 | 6.9 | 25323 | 24797 | 2.1 |
| **Slovenia**[d] | - | PAPI | - | - | - | | | | |
| **Switzerland**[e1] | - | CATI | - | - | - | | | | |
| **Switzerland** | 84.4, .47 | CAWI | 1.18 | .84 | .84 | 8.7 | | | |

***Note***. PCM = Rasch partial credit model, GPCM = generalized PCM, RelRed = relative reduction in -2*LL* from PCM to GPCM, *α* = Cronbach's alpha, PSI = Person Separation Index, *$p < .05$, **$p < .01$. The chi-square test for overall data-model fit using PCM was based on *G* = 8 groups of respondents (*df* = 7 for a single item and *df* = 84 for 12 items) and a reduced sample size with 20 persons for each of 36 thresholds *n* = 720: $\chi^2$(*df* = 84, *n* = 720), where number of thresholds = 12 x (4-1) = 36.

[a]Austria collected data using different modes (CATI and CAWI) in two comparable samples

[e2]Bulgaria applied CAPI (*n* = 402) and CAWI (*n* = 463) in small samples

[b]Czechia applied CATI in a medium sample *n* = 532 with 8 extreme scorers

[c]Israel applied CATI in a small sample *n* = 311 with 25 extreme scorers

[d]Slovenia applied PAPI in a minor sample *n* = 12, no analysis reported

[e1]Switzerland applied CATI in a minor sample *n* = 192, no analysis reported

[f]France, Hungary, Portugal, Russia and Slovenia (CAWI and CAPI) have acceptable overall fit to PCM when sample size is reduced to 10 persons per threshold (*n* = 360) for chi-square test. The GPCM was estimated for each of these five countries.

[g]Mean Rasch-based health literacy proficiency (using the PCM with 4-point raw score)

[h]Mean Rasch-based score when data for the two modes are merged to form a common point of zero. When analysed separately, the mean is 1.06 and 1.57 (Austria), .80 and 1.23 (Czechia), and .87 and 1.81 (Israel) for CAWI and CATI, respectively. Only the two Austrian samples are comparable.

**Table B1.** $HLS_{19}$-DIGI overall analyses

| Country | $\chi^2$, *p* | Mode | Mean | Reliability | | Dim (%) |
|---|---|---|---|---|---|---|
| | | | | *α* | *PSI* | |
| **Austria** | 18.3, .98 | CATI | 1.20 | .89 | .86 | 8.5 |
| **Austria** | 39.2, .18 | CAWI | .65 | .89 | .87 | 8.6 |
| **Belgium** | 56.1, .01* | CAWI | .14 | .91 | .89 | 10.6 |
| **Czechia** | 55.0, .01* | CAWI | .28 | .89 | .87 | 7.6 |
| **Denmark** | 63.9, .00** | CAWI | .94 | .92 | .90 | 8.3 |
| **Germany** | 16.5, .99 | PAPI | -.54 | .91 | .89 | 6.2 |
| **Hungary** | 80.3, 00** | CATI | .72 | .84 | .82 | 11.5 |
| **Ireland** | 38.8, .19 | CAWI | .64 | .86 | .83 | 5.5 |
| **Israel** | 97.0, .00** | CAWI | .57 | .88 | .87 | 8.5 |
| **Norway** | 40.1, .15 | CATI | 1.70 | .87 | .83 | 7.2 |
| **Portugal** | 106.1, 00** | CATI | 1.03 | .89 | .83 | 8.0 |
| **Switzerland** | 46.5, .05 | CAWI | 0.20 | .91 | .90 | 13.6 |

*$p$ < .05, **$p$ < .01. The chi-square test for overall data-model fit using PCM was based on a reduced sample size with 20 persons for each of 24 thresholds, $n$ = 480. Reducing sample size down to $n$ = 240 or 10 persons per threshold, also data from Denmark, Hungary, Israel and Portugal display acceptable overall data-model fit.

**Table C1a.** $HLS_{19}$-COM-Q11 overall analyses

| Country | $\chi^2$, *p* | Mode | Mean | Reliability | | Dim (%) |
|---|---|---|---|---|---|---|
| | | | | *α* | *PSI* | |
| **Austria** | 81.5, .00** | CATI | 2.57 | .91 | .86 | 6.1 |
| **Austria** | 138.1, 00** | CAWI | 1.97 | .93 | .89 | 7.5 |
| **Germany** | 84.5, .00** | PAPI | 1.38 | .90 | .89 | 7.9 |
| **Slovenia** | 108.1, 00** | CAWI | 2.73 | .94 | .88 | 9.9 |
| **Slovenia** | 94.2, .00** | CAPI | 2.55 | .94 | .88 | 4.8 |

*$p$ < .05, **$p$ < .01. The chi-square test for overall data-model fit using PCM was based on a reduced sample size with 20 persons for each of 33 thresholds, $n$ = 660. Reducing sample size down to $n$ = 330 or 10 persons improved fit and displayed acceptable overall data-model fit.

**Table C1b.** $HLS_{19}$-COM-Q6 overall analyses

| Country | $\chi^2$, *p* | Mode | Mean | Reliability | | Dim (%) |
|---|---|---|---|---|---|---|
| | | | | *α* | *PSI* | |
| **Austria** | 33.2, .1 | CATI | 2.39 | .86 | .75 | 5.3 |
| **Austria** | 53.5, 00** | CAWI | 1.76 | .89 | .82 | 3.9 |
| **Belgium** | 57.3, 00** | CAWI | 2.20 | .90 | .82 | 4.4 |
| **Bulgaria** | 62.9, 00** | CAPI | 1.34 | .87 | .80 | 6.5 |
| **Bulgaria** | 93.2, 00** | CAWI | 1.58 | .88 | .82 | 5.3 |
| **Czechia** | 51.3, .00** | CAWI | 1.54 | .88 | .83 | 5.1 |
| **Denmark** | 86.7, .00** | CAWI | 1.97 | .90 | .83 | 7.5 |
| **France** | 44.5, .01* | CAWI | 1.85 | .89 | .83 | 3.6 |
| **Germany** | 34.6, .07 | PAPI | 1.21 | .84 | .81 | 5.0 |
| **Hungary** | 52.1, .00** | CATI | 1.88 | .88 | .77 | 3.0 |
| **Slovenia** | 47.7, .00** | CAWI | 2.47 | .89 | .79 | 4.4 |
| **Slovenia** | 45.8, .00** | CAPI | 2.36 | .90 | .78 | 3.0 |

*$p$ < .05, **$p$ < .01. The chi-square test for overall data-model fit using PCM was based on a reduced sample size with 20 persons for each of 18 thresholds, $n$ = 360. Reducing sample size down to $n$ = 180 or 10 persons per threshold, also data from Belgium, Czechia, Denmark, Hungary and Slovenia display acceptable overall data-model fit.

**Table D1.** $HLS_{19}$-NAV overall analyses

| Country | $\chi^2$, $p$ | Mode | Mean | Reliability | | Dim (%) |
|---|---|---|---|---|---|---|
| | | | | $\alpha$ | *PSI* | |
| **Austria** | 59.3, .13 | CATI | .91 | .92 | .90 | 9.6 |
| **Austria** | 82.0, 00** | CAWI | .19 | .92 | .91 | 10.5 |
| **Belgium** | 107.3, 00** | CAWI | -.07 | .93 | .92 | 11.5 |
| **Czechia** | 73.4, .01* | CAWI | -.15 | .93 | .92 | 5.3 |
| **France** | 165.2, 00** | CAWI | .11 | .94 | .93 | 8.3 |
| **Germany** | 73.8, 01* | PAPI | -.31 | .88 | .88 | 12.2 |
| **Portugal** | 122,9, .00** | CATI | .21 | .94 | .88 | 9.3 |
| **Slovenia** | 137.5, 00** | CAWI | .96 | .94 | .92 | 9.2 |
| **Slovenia** | 105.0, .00** | CAPI | .63 | .93 | .91 | 10.3 |
| **Switzerland** | 74.3, 01* | CAWI | .04 | .92 | .91 | 9.7 |

*$p$ < .05, **$p$ < .01. The chi-square test for overall data-model fit using PCM was based on $G$ = 5 groups of respondents and a reduced sample size with 20 persons for each of 36 thresholds $n$ = 720. Reducing sample size down to $n$ = 360 or 10 persons per threshold also data collected in Belgium, Portugal and Slovenia display acceptable overall data-model fit (not France).
***Note***. Analysis of dimensionality was based on the two **theoretical** subdimensions (items OPNHL1–5 and OPNHL6–11(12)).

**Table E1.** $HLS_{19}$-VAC (4 items, q19, q22, q26, q29) overall analyses

| Country | $\chi^2$, $p$ | Mode | Mean | Reliability | | Dim (%)[a] |
|---|---|---|---|---|---|---|
| | | | | $\alpha$ | *PSI* | |
| **Austria** | 48.6, .00** | CATI | 1.69 | .82 | .68 | 8.9 |
| **Austria** | 26.5, .05 | CAWI | 1.49 | .83 | .72 | 9.1 |
| **Belgium** | 31.9, .01* | CAWI | 1,06 | .88 | .82 | 18.1 |
| **Bulgaria** | 27.4, .04* | CAPI | .16 | .72 | .67 | 7.8 |
| **Bulgaria** | 85.1, .00** | CAWI | .31 | .71 | .63 | 13.0 |
| **Czechia** | 18.3, .11 | CAWI | .87 | .80 | .73 | 15.8 |
| **Czechia** | 39.9, .00** | CATI | 1.60 | .76 | .55 | 11.7 |
| **Germany** | 24.8, .07 | PAPI | .90 | .76 | .71 | 9.6 |
| **Hungary** | 17.0, .15 | CATI | 1.38 | .78 | .70 | 11.6 |
| **Ireland** | 22.9, .12 | CAWI | 1.15 | .77 | .57 | 5.7 |
| **Italy** | 23.9, .05 | CATI | .89 | .75 | .64 | 6.8 |
| **Italy** | 10.7, .56 | CAWI | .98 | .81 | .73 | 5.8 |
| **Norway** | 36.2, .00** | CATI | 1.49 | .77 | .66 | 6.7 |
| **Portugal** | 74.2, .00** | CATI | 1.36 | .72 | .55 | 13.1 |
| **Slovenia** | 29.3, .02* | CAWI | 1.28 | .77 | .69 | 7.5 |
| **Slovenia** | 20.7, .06 | CAPI | 1.21 | .82 | .78 | 8.1 |

[a]Dimensionality (Dim) refers to the proportion of respondents with significantly different person estimates on $HLS_{19}$-Q12 and $HLS_{19}$-VAC,*$p$ < .05, **$p$ < .01. The chi-square test for overall data-model fit using PCM was based on $G$ = 5 groups of respondents and a sample size with **30** persons for each of 12 thresholds $n$ = 360. We used 30 persons and not 20 per threshold as there are few items in the VAC-scale. Reducing sample size down to $n$ = 240 or 20 persons per threshold also data collected in Norway and Czechia (CATI) display acceptable overall data-model fit. Reducing sample size down to $n$ = 120 or 10 persons per threshold data collected in Austria (CATI), Bulgaria (CAWI) and Portugal display acceptable fit.

# REFERENCES


Adams, R. J., Wu, M. L., Cloney, D., & Wilson, M. R. (2020). *ACER ConQuest: Generalised Item Response Modelling Software [Computer software] Version 5*. Camberwell, Victoria: Australian Council for Educational Research.

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

Andrich, D. (2011). *Advanced course in Rasch Measurement of Modern test Theory: Responses to the Model – Analysis of Residuals [Lecture note 8]. Unpublished manuscript*.

Andrich, D., & Marais, I. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. Springer.

Andrich, D., & Sheridan, B. (2019). *RUMM2030 Plus [Computer software]*. Rumm Laboratory Pty Ltd.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical Theories on Mental Test Scores, Reading: Mass./Addison-Wesley* (pp. 397–479). Reading.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *Bmj*, *310*(6973), 170.

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2008). The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. De Leeuw, L. Japec, & P. J. Lavrakas, *Advances in Telephone Survey Methodology* (pp. 250–275). Wiley-Blackwell. https://experts.nebraska.edu/en/publications/the-effects-of-mode-and-format-on-answers-to-scalar-questions-in-

de Ayala, R. J. (2009). *The theory and practice of item response theory.* Guilford Press.

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*(1), 1–17.

Guyer, R., & Thompson, N. A. (2014). *User's Manual for Xcalibre item response theory calibration software, version 4.2.2 and later*. Woodbury MN: Assessment Systems Corporation.

Lugtig, P., Lensvelt-Mulders, G. J., Frerichs, R., & Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. *International Journal of Market Research*, *53*(5), 669–686.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, *16*(2), 159–176.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests (Expanded ed.)*. University of Chicago Press.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement No. 17*. http://www.psychometrika.org/journal/online/MN17.pdf

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*(4), 567–577.

Van Wyke, J., & Andrich, D. (2006). A typology of polytomously scored mathematics items disclosed by the Rasch model: Implications for constructing a continuum of achievement. In: Report No 2 ARC Linkage Grant LP0454080. Maintaining Invariant Scales in State, National and International Level Assessments. Perth, Western Australia. Murdoch University. *Unpublished Report, Perth, Australia*.
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.9714&rep=rep1&type=pdf

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450. https://doi.org/10.1007/BF02294627

Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA.

Ye, C., Fulton, J., & Tourangeau, R. (2011). More positive or more extreme? A meta-analysis of mode differences in response choice. *Public Opinion Quarterly*, *75*(2), 349–365.

# APPENDIX

The appendix consists of tables with item-specific statistics for the different scales. We have based all analyses on the PCM.

**Table A2.** $HLS_{19}$-Q12 single item statistics

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| **Austria** | 4 | 0.005 | 2.419 | 0.933 | 1.03 | - |
| | 7 | -1.279 | 2.160 | 0.951 | 1.00 | - |
| | 10 | 2.459 | 17.211 | 0.016 | 1.08 | - |
| | 16 | -0.455 | 6.671 | 0.464 | 1.01 | - |
| | 18 | 0.013 | 5.833 | 0.559 | 1.03 | - |
| | 23 | -5.409 | 23.759 | 0.001 | 0.91 | - |
| | 24 | -1.678 | 2.714 | 0.910 | 1.00 | - |
| | 31 | 3.164 | 12.84 | 0.076 | 1.12 | - |
| | 32 | -3.603 | 9.121 | 0.244 | 0.95 | - |
| | 37 | 1.267 | 6.740 | 0.456 | 1.07 | - |
| | 42 | -2.954 | 8.949 | 0.256 | 0.98 | - |
| | 44 | -2.942 | 9.744 | 0.204 | 0.96 | - |
| | | | | | | |
| **Austria** | 4 | 0.712 | 13.334 | 0.064 | 1.03 | - |
| **CAWI** | 7 | -0.625 | 4.803 | 0.684 | 1.00 | - |
| | 10 | 2.018 | 23.709 | 0.001 | 1.09 | - |
| | 16 | -0.152 | 18.450 | 0.010 | 1.03 | health# |
| | 18 | 1.529 | 8.095 | 0.324 | 1.07 | - |
| | 23 | -2.100 | 12.958 | 0.073 | 0.96 | agedico* |
| | 24 | -1.987 | 8.540 | 0.287 | 0.97 | - |
| | 31 | 2.297 | 4.984 | 0.662 | 1.10 | soc.level |
| | 32 | -2.704 | 15.650 | 0.029 | 0.94 | - |
| | 37 | 2.704 | 27.834 | <0.001 | 1.14 | gender, agedico* |
| | 42 | -1.804 | 12.366 | 0.089 | 0.97 | - |
| | 44 | -3.236 | 22.541 | 0.002 | 0.92 | health* |
| | | | | | | |
| **Belgium** | 4 | 0.872 | 9.323 | 0.230 | 1.02 | - |
| n=997[a] | 7 | 1.320 | 8.215 | 0.314 | 1.06 | - |
| | 10 | -0.525 | 16.173 | 0.024 | 0.98 | - |
| | 16 | -1.455 | 10.353 | 0.170 | 0.98 | - |
| | 18 | 1.626 | 6.309 | 0.504 | 1.08 | - |
| | 23 | -2.795 | 17.460 | 0.015 | 0.91 | agedico*, agecat1*, agecat2*, employment* |
| | 24 | -0.943 | 12.120 | 0.097 | 0.97 | - |
| | 31 | 3.956 | 26.746 | <0.001 | 1.18 | agecat1, employment* |
| | 32 | -2.828 | 12.580 | 0.083 | 0.91 | - |
| | 37 | 0.985 | 6.881 | 0.441 | 1.05 | gender, pay bills* |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| | 42 | -1.221 | 9.097 | 0.246 | 0.94 | - |
| | 44 | -1.711 | 16.165 | 0.024 | 0.95 | health* |
| | | | | | | |
| **Bulgaria** | 4 | -1249 | 4.004 | 0.779 | 0.93 | - |
| **CAPI** | 7 | -1.182 | 10.612 | 0.156 | 0.94 | gender* |
| **n=402** | 10 | -1.013 | 9.802 | 0.200 | 0.91 | - |
| | 16 | -1.061 | 9.249 | 0.235 | 0.95 | - |
| | 18 | -1.197 | 13.424 | 0.062 | 0.94 | - |
| | 23 | 4.618 | 38.920 | <0.001 | 1.30 | soc.level*, health* |
| | 24 | -1.984 | 14.623 | 0.041 | 0.94 | - |
| | 31 | 1.172 | 4.447 | 0.727 | 1.05 | - |
| | 32 | -1.232 | 10.509 | 0.162 | 0.94 | agecat1*, employment* |
| | 37 | 0.235 | 6.067 | 0.532 | 1.04 | - |
| | 42 | -0.062 | 8.384 | 0.300 | 1.05 | - |
| | 44 | 1.071 | 11.121 | 0.133 | 1.08 | agecat2* |
| | | | | | | |
| **Bulgaria** | 4 | 0.963 | 11.540 | 0.117 | 1.07 | - |
| **CAWI** | 7 | -0.050 | 3.022 | 0.883 | 0.97 | - |
| n=463 | 10 | -2.814 | 19.315 | 0.007 | 0.83 | - |
| | 16 | -1.914 | 9.225 | 0.237 | 0.89 | - |
| | 18 | -0.658 | 3.524 | 0.833 | 0.98 | - |
| | 23 | 9.427 | 131.899 | <0.001 | 1.61 | health* |
| | 24 | -2.274 | 15.227 | 0.033 | 0.88 | agecat1*, health*# |
| | 31 | 0.058 | 10.379 | 0.168 | 1.00 | - |
| | 32 | -1.374 | 6.253 | 0.511 | 0.95 | - |
| | 37 | -1.499 | 11.318 | 0.125 | 0.89 | - |
| | 42 | 0.376 | 2.718 | 0.910 | 1.02 | - |
| | 44 | 0.455 | 13.726 | 0.056 | 0.98 | agedico*, agecat2* |
| | | | | | | |
| **Czechia** | 4 | -2.090 | 11.767 | 0.109 | 0.93 | - |
| **CATI** | 7 | -2.531 | 15.554 | 0.030 | 0.89 | - |
| n=530 | 10 | 1.116 | 4.264 | 0.749 | 1.04 | - |
| | 16 | 0.093 | 13.059 | 0.071 | 1.08 | agecat1*, health*# |
| | 18 | -1.199 | 15.338 | 0.032 | 0.98 | - |
| | 23 | -3.074 | 10.932 | 0.142 | 0.90 | - |
| | 24 | -0.049 | 5.353 | 0.617 | 1.07 | - |
| | 31 | -0.108 | 7.300 | 0.398 | 1.03 | agedico*, agecat2* |
| | 32 | -1.270 | 5.762 | 0.568 | 0.99 | agedico*, agecat1*, agecat2*, education*, employment* |
| | 37 | 0.228 | 6.248 | 0.511 | 1.13 | - |
| | 42 | -0.423 | 2.792 | 0.904 | 1.02 | - |
| | 44 | 1.845 | 11.938 | 0.103 | 1.13 | health* |
| | | | | | | |
| **Czechia** | 4 | -2.482 | 15.614 | 0.029 | 0.93 | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| **CAWI** | 7 | -1.674 | 8.617 | 0.281 | 0.95 | agedico |
| | 10 | -0.506 | 26.912 | <0.001 | 0.98 | education |
| | 16 | -0.674 | 9.357 | 0.228 | 1.01 | - |
| | 18 | -0.980 | 17.101 | 0.017 | 1.00 | - |
| | 23 | -1.014 | 2.523 | 0.925 | 1.02 | gender, agedico, agecat1, agecat2 |
| | 24 | -1.200 | 10.266 | 0.174 | 1.00 | - |
| | 31 | -1.275 | 10.258 | 0.174 | 0.97 | - |
| | 32 | 0.115 | 20.740 | 0.004 | 1.06 | agedico, education*, agecat2 |
| | 37 | 0.877 | 10.932 | 0.142 | 1.07 | - |
| | 42 | -2.059 | 12.450 | 0.087 | 0.95 | - |
| | 44 | 1.415 | 13.769 | 0.055 | 1.10 | - |
| | | | | | | |
| **Denmark** | 4 | -0.761 | 2.022 | 0.959 | 1.00 | - |
| | 7 | 0.239 | 4.3 | 0.745 | 1.04 | - |
| | 10 | -0.640 | 17.394 | 0.015 | 0.97 | - |
| | 16 | -4.320 | 13.206 | 0.067 | 0.91 | - |
| | 18 | 0.325 | 4.713 | 0.695 | 1.03 | - |
| | 23 | -3.011 | 9.08 | 0.247 | 0.97 | gender, agedico*, agecat2* |
| | 24 | -2.366 | 10.36 | 0.169 | 0.98 | - |
| | 31 | 2.057 | 11.196 | 0.130 | 1.05 | - |
| | 32 | -6.151 | 28.741 | <0.001 | 0.89 | gender |
| | 37 | 1.162 | 7.642 | 0.365 | 1.07 | - |
| | 42 | -0.794 | 3.572 | 0.828 | 1.02 | agedico*, agecat1*, agecat2*, employment* |
| | 44 | 3.862 | 7.698 | 0.360 | 1.10 | - |
| | | | | | | |
| **France** | 4 | 3.636 | 32.900 | <0.001* | 1.16 | - |
| | 7 | -1.091 | 5.880 | 0.554 | 1.01 | - |
| | 10 | -0.140 | 24.333 | 0.001 | 1.01 | - |
| | 16 | -0.376 | 56.743 | <0.001* | 1.07 | - |
| | 18 | -1.658 | 25.695 | 0.001 | 0.96 | - |
| | 23 | -4.237 | 9.312 | 0.231 | 0.94 | agedico*, agecat1*, agecat2*, employment* |
| | 24 | -5.807 | 17.438 | 0.015 | 0.91 | - |
| | 31 | 0.464 | 32.524 | <0.001* | 1.03 | - |
| | 32 | -2.572 | 18.651 | 0.009 | 0.96 | - |
| | 37 | 0.745 | 16.855 | 0.018 | 1.12 | employment |
| | 42 | -4.608 | 11.991 | 0.101 | 0.93 | agecat1, pay bills |
| | 44 | -4.150 | 11.803 | 0.107 | 0.94 | - |
| | | | | | | |
| **Germany** | 4 | 0.934 | 2.859 | 0.898 | 1.03 | - |
| | 7 | -0.738 | 7.782 | 0.352 | 0.97 | - |
| | 10 | -1.034 | 7.162 | 0.412 | 0.95 | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | | DIF *n*=1080 |
|---|---|---|---|---|---|---|---|
| | 16 | 0.122 | 3.922 | 0.789 | 1.03 | | - |
| | 18 | -0.644 | 3.425 | 0.843 | 0.98 | | - |
| | 23 | -2.663 | 17.032 | 0.017 | 0.95 | | agedico |
| | 24 | -1.446 | 3.805 | 0.802 | 0.98 | | - |
| | 31 | 0.323 | 5.862 | 0.556 | 1.00 | | - |
| | 32 | -2.359 | 10.757 | 0.150 | 0.95 | | - |
| | 37 | 4.039 | 43.221 | <0.001* | 1.15 | | education, soc.level |
| | 42 | 1.084 | 5.693 | 0.576 | 1.05 | | - |
| | 44 | 0.602 | 3.015 | 0.884 | 1.04 | | - |
| | | | | | | | |
| **Hungary** | 4 | 2.381 | 29.997 | <0.001 | 1.18 | | - |
| | 7 | -2.105 | 5.519 | 0.597 | 1.01 | | education |
| | 10 | -2.606 | 17.867 | 0.013 | 0.94 | | - |
| | 16 | -3.711 | 22.214 | 0.002 | 0.92 | | - |
| | 18 | 1.488 | 22.086 | 0.003 | 1.13 | | - |
| | 23 | -4.287 | 19.474 | 0.007 | 0.90 | | education |
| | 24 | -3.247 | 5.937 | 0.547 | 0.95 | | - |
| | 31 | 3.268 | 45.190 | <0.001* | 1.19 | | education |
| | 32 | -3.804 | 10.828 | 0.146293 | 0.91 | | agecat1, employment* |
| | 37 | -2.513 | 10.178 | 0.178704 | 0.98 | | - |
| | 42 | -3.759 | 8.902 | 0.259741 | 0.94 | | - |
| | 44 | -2.278 | 7.328 | 0.395593 | 1.00 | | - |
| | | | | | | | |
| **Ireland** | | | | | Set A | Set B | |
| N > 3000 | 4 | 1.837 | 3.999 | 0.780 | 1.08 | 1.05 | - |
| | 7 | 3.400 | 2.243 | 0.945 | 1.06 | 1.05 | agecat1 |
| | 10 | 4.343 | 6.033 | 0.536 | 1.08 | 1.04 | - |
| | 16 | -4.564 | 20.354 | 0.005 | 0.94 | 0.94 | - |
| | 18 | 3.351 | 4.110 | 0.767 | 1.02 | 1.10 | - |
| | 23 | -5.584 | 24.935 | 0.001 | 0.88 | 0.93 | - |
| | 24 | -1.694 | 6.155 | 0.522 | 0.98 | 0.98 | - |
| | 31 | 11.082 | 42.714 | <0.001* | 1.19 | 1.22 | - |
| | 32 | -4.466 | 18.342 | 0.011 | 0.96 | 0.95 | - |
| | 37 | -1.004 | 13.969 | 0.052 | 0.97 | 0.98 | - |
| | 42 | -0.410 | 8.937 | 0.257 | 1.04 | 1.02 | agedico, agecat1, agecat2 |
| | 44 | -3.061 | 18.514 | 0.010 | 0.94 | 0.95 | - |
| | | | | | | | |
| **Israel** | 4 | -0.130 | 4.163 | 0.761 | 0.99 | | - |
| **CAWI** | 7 | 0.818 | 5.447 | 0.606 | 1.03 | | - |
| | 10 | 0.072 | 19.638 | 0.006 | 0.99 | | agecat1 |
| | 16 | -0.930 | 16.340 | 0.022 | 0.95 | | - |
| | 18 | -0.998 | 7.814 | 0.349 | 0.97 | | - |
| | 23 | -2.142 | 13.312 | 0.065 | 0.95 | | agedico, agecat2 |
| | 24 | -0.365 | 5.276 | 0.626 | 1.05 | | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| | 31 | 0.048 | 8.687 | 0.276 | 0.98 | - |
| | 32 | -0.621 | 22.741 | 0.002 | 1.03 | gender |
| | 37 | 0.022 | 5.011 | 0.659 | 1.03 | - |
| | 42 | 0.211 | 16.057 | 0.025 | 1.03 | - |
| | 44 | 0.350 | 9.553 | 0.215 | 1.04 | - |
| | | | | | | |
| **Italy** | 4 | -1.829 | 2.629 | 0.917 | 1.02 | - |
| **CATI** | 7 | -3.304 | 16.510 | 0.021 | 0.90 | - |
| n=551 | 10 | -1.414 | 18.830 | 0.009 | 1.00 | - |
| | 16 | -1.197 | 19.147 | 0.008 | 1.02 | - |
| | 18 | -0.854 | 5.853 | 0.557 | 1.05 | - |
| | 23 | -3.488 | 16.136 | 0.024 | 0.88 | - |
| | 24 | -3.198 | 14.797 | 0.039 | 0.89 | - |
| | 31 | 1.194 | 16.344 | 0.022 | 1.13 | - |
| | 32 | -2.323 | 12.558 | 0.084 | 0.99 | agecat1* |
| | 37 | -0.252 | 20.042 | 0.006 | 1.12 | - |
| | 42 | -2.923 | 2.917 | 0.893 | 0.95 | - |
| | 44 | -0.745 | 13.954 | 0.052 | 1.09 | - |
| | | | | | | |
| **Italy** | 4 | 2.496 | 20.916 | 0.004 | 1.09 | - |
| **CAWI** | 7 | -4.491 | 34.767 | <0.001 | 0.90 | - |
| | 10 | -0.892 | 12.313 | 0.091 | 0.99 | - |
| | 16 | -0.853 | 28.921 | <0.001 | 1.02 | - |
| | 18 | -0.344 | 21.326 | 0.003 | 1.02 | - |
| | 23 | -4.380 | 19.852 | 0.006 | 0.93 | - |
| | 24 | -3.415 | 32.903 | <0.001 | 0.95 | - |
| | 31 | 1.040 | 39.284 | <0.001 | 1.06 | - |
| | 32 | 1.090 | 27.021 | <0.001 | 1.07 | - |
| | 37 | -0.279 | 9.901 | 0.194 | 1.03 | - |
| | 42 | -0.943 | 5.502 | 0.599 | 1.00 | - |
| | 44 | -2.223 | 11.884 | 0.105 | 0.97 | - |
| | | | | | | |
| **Norway** | 4 | 1.057 | 3.115 | 0.874 | 1.03 | - |
| | 7 | 0.777 | 4.306 | 0.744 | 1.02 | - |
| | 10 | 3.180 | 10.700 | 0.152 | 1.08 | - |
| | 16 | -4.522 | 18.621 | 0.009 | 0.89 | - |
| | 18 | 1.335 | 5.501 | 0.599 | 1.05 | - |
| | 23 | -5.696 | 28.421 | <0.001 | 0.89 | - |
| | 24 | -2.661 | 9.317 | 0.231 | 0.95 | - |
| | 31 | 6.412 | 28.895 | <0.001 | 1.17 | - |
| | 32 | -3.400 | 13.787 | 0.055 | 0.95 | - |
| | 37 | -1.220 | 5.010 | 0.659 | 0.98 | - |
| | 42 | 0.440 | 6.405 | 0.493 | 1.04 | agedico, agecat2, employment |
| | 44 | 0.593 | 3.235 | 0.862 | 1.04 | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | | DIF *n*=1080 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Portugal** | 4 | -4.795 | 11.852 | 0.106 | 1.02 | | - |
| | 7 | -3.945 | 8.943 | 0.257 | 1.07 | | soc.level, employment |
| | 10 | -2.447 | 71.880 | <0.001* | 1.07 | | soc. level* |
| | 16 | -2.631 | 84.219 | <0.001* | 1.02 | | pay bills[#] |
| | 18 | -2.901 | 29.074 | <0.001 | 1.05 | | - |
| | 23 | -5.229 | 11.061 | 0.136 | 0.91 | | - |
| | 24 | -6.852 | 8.467 | 0.293 | 0.84 | | - |
| | 31 | -1.955 | 59.742 | <0.001* | 1.06 | | - |
| | 32 | -8.566 | 26.022 | 0.001 | 0.79 | | agedico*[#], agecat1*, agecat2* |
| | 37 | -4.769 | 17.228 | 0.016 | 0.87 | | - |
| | 42 | -4.668 | 5.109 | 0.647 | 0.88 | | - |
| | 44 | -3.794 | 5.104 | 0.647 | 1.04 | | agecat2 |
| | | | | | | | |
| **Russia** | | | | | Set A | Set B | |
| | 4 | -4.912 | 14.363 | 0.045 | 1.06 | 1.07 | - |
| | 7 | -7.570 | 15.121 | 0.035 | 0.98 | 0.98 | - |
| | 10 | -5.882 | 55.904 | <0.001* | 1.01 | 0.94 | - |
| | 16 | -7.664 | 17.821 | 0.013 | 0.98 | 1.10 | - |
| | 18 | -8.720 | 3.384 | 0.847 | 0.97 | 1.03 | - |
| | 23 | -9.794 | 14.145 | 0.049 | 0.93 | 0.95 | - |
| | 24 | -11.843 | 3.327 | 0.853 | 0.92 | 1.10 | - |
| | 31 | -7.132 | 14.769 | 0.039 | 1.03 | 0.97 | - |
| | 32 | -10.220 | 38.239 | <0.001* | 0.93 | 0.97 | - |
| | 37 | -4.727 | 13.567 | 0.059 | 1.10 | 1.06 | - |
| | 42 | -7.551 | 3.730 | 0.810 | 0.96 | 1.03 | - |
| | 44 | -2.978 | 8.516 | 0.289 | 1.09 | 1.09 | - |
| | | | | | | | |
| **Slovakia** | 4 | -1.142 | 4.134 | 0.764 | 0.99 | | - |
| | 7 | -2.017 | 14.02 | 0.051 | 0.96 | | - |
| | 10 | -0.383 | 8.31 | 0.306 | 0.99 | | - |
| | 16 | 0.442 | 13.918 | 0.053 | 1.05 | | - |
| | 18 | -2.415 | 9.383 | 0.226 | 0.94 | | - |
| | 23 | -3.325 | 5.049 | 0.654 | 0.95 | | - |
| | 24 | -1.596 | 4.807 | 0.684 | 1.00 | | - |
| | 31 | 1.871 | 7.554 | 0.374 | 1.08 | | - |
| | 32 | -5.334 | 28.07 | <0.001 | 0.89 | | - |
| | 37 | 0.856 | 14.677 | 0.040 | 1.06 | | - |
| | 42 | -1.066 | 2.989 | 0.886 | 1.01 | | - |
| | 44 | 1.567 | 8.956 | 0.256 | 1.10 | | - |
| | | | | | | | |
| **Slovenia** | 4 | -1.2 | 2.466 | 0.930 | 0.97 | | - |
| **CAWI** | 7 | -3.031 | 12.854 | 0.076 | 0.91 | | agedico*, agecat2 |
| | 10 | 1.624 | 10.812 | 0.147 | 1.05 | | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| | 16 | -2.47 | 14.815 | 0.039 | 0.93 | - |
| | 18 | -3.293 | 18.822 | 0.009 | 0.91 | - |
| | 23 | -1.086 | 21.182 | 0.004 | 1.02 | agedico*, agecat1*, agecat2*, employment* |
| | 24 | -5.265 | 29.706 | <0.001 | 0.87 | - |
| | 31 | 7.859 | 99.689 | <0.001* | 1.32 | education*, pay bills |
| | 32 | -2.734 | 27.45 | <0.001 | 0.94 | education*, soc. level |
| | 37 | -1.791 | 26.717 | <0.001 | 0.96 | - |
| | 42 | -1.586 | 11.398 | 0.122 | 0.97 | - |
| | 44 | 3.574 | 26.423 | <0.001 | 1.17 | - |
| | | | | | | |
| **Slovenia** | 4 | -2.07 | 5.666 | 0.579 | 1.10 | - |
| **CAPI** | 7 | -5.838 | 7.645 | 0.365 | 0.93 | - |
| | 10 | -1.017 | 42.188 | <0.001* | 1.07 | - |
| | 16 | -5.588 | 22.078 | 0.003 | 0.94 | - |
| | 18 | -6.15 | 13.248 | 0.066 | 0.93 | - |
| | 23 | -3.874 | 9.672 | 0.208 | 1.01 | gender, agedico*, agecat1*, agecat2* |
| | 24 | -9.418 | 22.354 | 0.002 | 0.85 | - |
| | 31 | 3.795 | 56.132 | <0.001* | 1.25 | health |
| | 32 | -5.619 | 17.042 | 0.017 | 0.92 | agedico, agecat1, agecat2, employment |
| | 37 | -2.824 | 6.972 | 0.432 | 1.07 | - |
| | 42 | -4.697 | 8.398 | 0.299 | 0.96 | - |
| | 44 | -2.712 | 6.454 | 0.488 | 1.08 | - |
| | | | | | | |
| **Switzerland** | 4 | 2.891 | 11.241 | 0.128 | 1.06 | - |
| **CAWI** | 7 | -2.762 | 13.594 | 0.059 | 0.93 | - |
| | 10 | -0.039 | 4.479 | 0.723 | 1.00 | - |
| | 16 | -1.773 | 11.139 | 0.133 | 0.97 | - |
| | 18 | -3.141 | 12.152 | 0.096 | 0.94 | - |
| | 23 | -2.238 | 3.662 | 0.818 | 0.98 | agedico*, agecat1, agecat2* |
| | 24 | -0.713 | 4.188 | 0.758 | 1.02 | - |
| | 31 | 4.839 | 26.514 | <0.001 | 1.16 | pay bills |
| | 32 | -2.292 | 10.475 | 0.163 | 0.97 | - |
| | 37 | 2.155 | 16.058 | 0.025 | 1.11 | agedico*, agecat1, agecat2*, employment |
| | 42 | -3.206 | 6.710 | 0.460 | 0.97 | - |
| | 44 | -2.117 | 6.419 | 0.492 | 0.97 | - |

[+] not in complete data

[#] non-uniform DIF

*significant when sample size = 720

[a]sample size available for Rasch analysis (excluding extreme scorers) when sample size is less than 1000

Not reported in table: $HLS_{19}$-Q12 item 16 displayed unordered thresholds in the Belgian, Czech (CATI), Irish, and Norwegian data (not significant in the Norwegian data). In Czech (CATI) data also $HLS_{19}$-Q12 item 32 displayed unordered thresholds. In the Austrian data, we observed *in*significant unordered thresholds for $HLS_{19}$-Q12 item 4.

**Table B2.** $HLS_{19}$-DIGI (8 items version) single item statistics

| Country | Item | Fit.res. | Chi sq *n*=720 | Chi sq *p* | Infit MNSQ | DIF *n*=720 |
|---|---|---|---|---|---|---|
| **Austria** | OPDHL21 | -0.572 | 2.001 | 0.736 | 1.03 | - |
| | OPDHL22 | 0.303 | 2.469 | 0.650 | 1.04 | - |
| | OPDHL23 | -0.646 | 1.669 | 0.796 | 1.02 | - |
| | OPDHL24 | -1.58 | 2.897 | 0.575 | 0.98 | - |
| | OPDHL25 | 3.423 | 8.800 | 0.066 | 1.13 | - |
| | OPDHL26 | 0.478 | 0.553 | 0.968 | 1.05 | - |
| | OPDHL27 | -3.647 | 5.072 | 0.280 | 0.93 | - |
| | OPDHL28 | -2.193 | 3.964 | 0.411 | 0.96 | - |
| | | | | | | |
| **Austria** | OPDHL21 | 1.544 | 11.472 | 0.022 | 1.09 | - |
| **CAWI** | OPDHL22 | 0.077 | 8.856 | 0.065 | 0.99 | - |
| | OPDHL23 | -1.234 | 4.327 | 0.364 | 0.95 | education |
| | OPDHL24 | -0.637 | 11.352 | 0.023 | 0.98 | - |
| | OPDHL25 | 1.086 | 11.740 | 0.019 | 1.08 | - |
| | OPDHL26 | 0.517 | 14.502 | 0.006 | 1.08 | - |
| | OPDHL27 | -3.834 | 18.723 | 0.001 | 0.88 | - |
| | OPDHL28 | -0.698 | 7.147 | 0.128 | 0.99 | - |
| | | | | | | |
| **Belgium** | OPDHL21 | 0.854 | 7.852 | 0.097 | 1.14 | agecat1, employment |
| | OPDHL22 | -3.054 | 8.077 | 0.089 | 0.89 | agedico, agecat1, agecat2, employment |
| | OPDHL23 | -2.405 | 1.903 | 0.754 | 0.96 | - |
| | OPDHL24 | -0.452 | 6.662 | 0.155 | 1.03 | - |
| | OPDHL25 | 5.925 | 27.452 | <0.001* | 1.34 | agecat1 |
| | OPDHL26 | -1.531 | 1.896 | 0.755 | 1.03 | - |
| | OPDHL27 | -5.793 | 20.37 | <0.001 | 0.81 | - |
| | OPDHL28 | -2.608 | 9.968 | 0.041 | 0.91 | - |
| | | | | | | |
| **Czechia** | OPDHL21 | 1.549 | 25.159 | <0.001* | 1.12 | - |
| CAWI | OPDHL22 | -1.49 | 2.102 | 0.717 | 0.96 | agecat1 |
| | OPDHL23 | -3.252 | 10.869 | 0.0287 | 0.88 | - |
| | OPDHL24 | -1.518 | 10.717 | 0.030 | 0.94 | - |
| | OPDHL25 | 3.926 | 9.112 | 0.058 | 1.24 | - |
| | OPDHL26 | -0.584 | 9.643 | 0.047 | 1.03 | - |
| | OPDHL27 | -2.368 | 6.176 | 0.186 | 0.93 | - |
| | OPDHL28 | -1.267 | 8.775 | 0.067 | 0.95 | - |
| | | | | | | |
| **Denmark** | OPDHL21 | -3.504 | 7.577 | 0.108 | 1.01 | - |
| | OPDHL22 | -3.715 | 6.11 | 0.191 | 1 | - |

| Country | Item | Fit.res. | Chi sq *n*=720 | Chi sq *p* | Infit MNSQ | | DIF *n*=720 |
|---|---|---|---|---|---|---|---|
| | OPDHL23 | -6.656 | 12.999 | 0.011 | 0.93 | | - |
| | OPDHL24 | -2.667 | 7.17 | 0.127 | 1.03 | | - |
| | OPDHL25 | 4.816 | 41.999 | <0.001* | 1.23 | | - |
| | OPDHL26 | -2.379 | 5.121 | 0.275 | 1.05 | | agedico, agecat1*, agecat2, employment |
| | OPDHL27 | -9.606 | 8.841 | 0.065 | 0.86 | | - |
| | OPDHL28 | -6.122 | 5.996 | 0.120 | 0.91 | | - |
| | | | | | | | |
| **Germany** | OPDHL21 | 0.165 | 0.53 | 0.971 | 1.04 | | - |
| | OPDHL22 | -1.679 | 3.083 | 0.544 | 0.95 | | - |
| | OPDHL23 | -0.688 | 2.536 | 0.638 | 0.94 | | - |
| | OPDHL24 | -0.632 | 6.682 | 0.154 | 0.99 | | - |
| | OPDHL25 | 3.189 | 5.998 | 0.199 | 1.14 | | - |
| | OPDHL26 | -0.616 | 2.069 | 0.723 | 1.02 | | - |
| | OPDHL27 | -1.669 | 3.051 | 0.549 | 0.95 | | - |
| | OPDHL28 | 0.076 | 0.764 | 0.943 | 1.01 | | - |
| | | | | | | | |
| **Hungary** | OPDHL21 | -2.192 | 13.684 | 0.008 | 1.01 | | agedico, education |
| | OPDHL22 | -3.549 | 4.599 | 0.331 | 0.95 | | - |
| | OPDHL23 | -3.072 | 13.085 | 0.011 | 0.95 | | - |
| | OPDHL24 | 0.279 | 36.973 | <0.001* | 1.04 | | - |
| | OPDHL25 | 1.234 | 27.713 | <0.001* | 1.13 | | - |
| | OPDHL26 | -1.642 | 5.681 | 0.224 | 0.99 | | - |
| | OPDHL27 | -3.216 | 9.64 | 0.047 | 0.95 | | - |
| | OPDHL28 | -2.823 | 9.059 | 0.060 | 0.93 | | - |
| | | | | | | | |
| **Ireland** n > 3000 | | | | | Set A | Set B | |
| | OPDHL21 | -0.552 | 3.236 | 0.519 | 1.04 | 0.98 | - |
| | OPDHL22 | -1.564 | 6.147 | 0.188 | 0.91 | 0.96 | - |
| | OPDHL23 | -2.365 | 10.148 | 0.038 | 0.9 | 0.94 | - |
| | OPDHL24 | -0.501 | 2.188 | 0.701 | 0.96 | 0.98 | - |
| | OPDHL25 | 10.92 | 18.052 | 0.001* | 1.25 | 1.20 | - |
| | OPDHL26 | 3.169 | 1.591 | 0.811 | 1.05 | 1.10 | agecat2 |
| | OPDHL27 | -3.998 | 14.869 | 0.005* | 0.89 | 0.91 | - |
| | OPDHL28 | 2.956 | 2.035 | 0.729 | 1.08 | 1.03 | - |
| | | | | | | | |
| **Israel** | OPDHL21 | 1.398 | 10.367 | 0.035 | 1.11 | | - |
| CAWI | OPDHL22 | -1.316 | 3.833 | 0.429 | 0.96 | | - |
| | OPDHL23 | -1.046 | 3.586 | 0.465 | 0.97 | | - |
| | OPDHL24 | -1.583 | 10.596 | 0.032 | 0.95 | | - |
| | OPDHL25 | 3.252 | 23 | <0.001 | 1.21 | | - |
| | OPDHL26 | -1.442 | 5.828 | 0.212 | 1.01 | | - |
| | OPDHL27 | -2.81 | 11.232 | 0.024 | 0.92 | | - |
| | OPDHL28 | -1.935 | 4.187 | 0.381 | 0.94 | | - |
| | | | | | | | |

| Country | Item | Fit.res. | Chi sq *n*=720 | Chi sq *p* | Infit MNSQ | DIF *n*=720 |
|---|---|---|---|---|---|---|
| **Norway** | OPDHL21 | -0.997 | 5.31 | 0.257 | 1.01 | - |
| | OPDHL22 | -0.475 | 3.907 | 0.419 | 0.99 | - |
| | OPDHL23 | -5.126 | 12.361 | 0.015 | 0.89 | - |
| | OPDHL24 | -1.699 | 2.835 | 0.586 | 0.97 | - |
| | OPDHL25 | 3.5 | 6.339 | 0.175 | 1.16 | - |
| | OPDHL26 | 1.841 | 16.775 | 0.002 | 1.13 | agedico, agecat1, agecat2 |
| | OPDHL27 | -3.105 | 7.84 | 0.098 | 0.92 | - |
| | OPDHL28 | 0.199 | 4.784 | 0.310 | 1.02 | - |
| | | | | | | |
| **Portugal** | OPDHL21 | -3.721 | 25.026 | <0.001* | 0.97 | - |
| | OPDHL22 | -3.973 | 2.906 | 0.574 | 0.99 | - |
| | OPDHL23 | -5.64 | 22.039 | <0.001* | 0.86 | - |
| | OPDHL24 | -5.371 | 23.089 | <0.001* | 0.92 | - |
| | OPDHL25 | -2.775 | 5.689 | 0.224 | 1.09 | - |
| | OPDHL26 | -1.937 | 30.796 | <0.001* | 1.22 | - |
| | OPDHL27 | -5.47 | 19.947 | <0.001 | 0.84 | education[#] |
| | OPDHL28 | -0.925 | 29.589 | <0.001* | 1.14 | gender, soc.level |
| | | | | | | |
| **Switzerland** | OPDHL21 | 1.048 | 15.117 | 0.004 | 1.11 | - |
| **CAWI** | OPDHL22 | -2.262 | 4.18 | 0.382 | 1.00 | - |
| | OPDHL23 | -3.391 | 4.925 | 0.295 | 0.96 | - |
| | OPDHL24 | -3.594 | 10.805 | 0.029 | 0.95 | - |
| | OPDHL25 | 4.423 | 14.448 | 0.006 | 1.20 | gender |
| | OPDHL26 | -1.079 | 6.584 | 0.160 | 1.06 | agedico |
| | OPDHL27 | -4.535 | 7.186 | 0.126 | 0.91 | - |
| | OPDHL28 | -3.402 | 6.552 | 0.161 | 0.95 | - |

[#] non-uniform DIF

*significant when sample size = 480

Not reported in table: Response dependency was observed between items OPDHL21 and OPDHL22 in Danish (r = .31) and Hungarian (r = .32) data.

**Table C2a.** $HLS_{19}$-COM-Q11 single item statistics

| Country | Item | Fit.res. | Chi sq *n*=990 | Chi sq *p* | Infit MNSQ | DIF *n*=990 |
|---|---|---|---|---|---|---|
| **Austria** | OPCOM1 | 1.14 | 20.956 | <0.001* | 1.07 | - |
| | OPCOM2 | -1.807 | 1.936 | 0.748 | 1.02 | - |
| | OPCOM3 | -4.478 | 9.510 | 0.050 | 0.95 | - |
| | OPCOM4 | 3.868 | 19.668 | 0.001 | 1.21 | agedico, agecat2 |
| | OPCOM5 | -7.175 | 9.744 | 0.045 | 0.89 | - |
| | OPCOM6 | -6.666 | 14.962 | 0.005 | 0.90 | - |
| | OPCOM7 | 3.209 | 14.114 | 0.007 | 1.18 | - |
| | OPCOM8 | -6.835 | 13.580 | 0.009 | 0.91 | - |

| Country | Item | Fit.res. | Chi sq *n*=990 | Chi sq *p* | Infit MNSQ | DIF *n*=990 |
|---|---|---|---|---|---|---|
| | OPCOM9 | -5.135 | 7.025 | 0.135 | 0.92 | - |
| | OPCOM10 | 0.23 | 6.174 | 0.187 | 1.12 | - |
| | OPCOM11 | -1.219 | 4.636 | 0.327 | 1.04 | - |
| | | | | | | |
| **Austria** | OPCOM1 | 1,287 | 18,509 | 0.001 | 1.17 | - |
| **CAWI** | OPCOM2 | -3,040 | 10,878 | 0,028 | 0.99 | - |
| | OPCOM3 | -3,632 | 15,986 | 0.003 | 1.01 | - |
| | OPCOM4 | -0,314 | 11,861 | 0.018 | 1.12 | - |
| | OPCOM5 | -4,925 | 12,92 | 0.012 | 0.94 | gender* |
| | OPCOM6 | -5,384 | 16,863 | 0.002 | 0.90 | - |
| | OPCOM7 | 4,736 | 31,599 | <0.001* | 1.38 | gender*, agedico, education |
| | OPCOM8 | -6,026 | 22,093 | <0.001 | 0.94 | - |
| | OPCOM9 | -4,990 | 25,019 | <0.001* | 0.92 | - |
| | OPCOM10 | 4,009 | 39,136 | <0.001* | 1.38 | - |
| | OPCOM11 | -0,814 | 2,275 | 0.685 | 1.11 | - |
| | | | | | | |
| **Germany** | OPCOM1 | -1.152 | 31,491 | <0.001* | 1.05 | - |
| | OPCOM2 | -2.428 | 2,474 | 0.649 | 1 | - |
| | OPCOM3 | -4.387 | 21,444 | <0.001 | 0.93 | - |
| | OPCOM4 | 2.312 | 9,542 | 0,049 | 1.11 | - |
| | OPCOM5 | -3.771 | 5,285 | 0,259 | 0.96 | - |
| | OPCOM6 | -5.928 | 8,83 | 0,066 | 0.86 | - |
| | OPCOM7 | 2.126 | 12,039 | 0,017 | 1.1 | education* |
| | OPCOM8 | -6.269 | 11,362 | 0,023 | 0.89 | - |
| | OPCOM9 | -2.298 | 6,072 | 0,194 | 0.97 | - |
| | OPCOM10 | 2.169 | 11,987 | 0,018 | 1.15 | - |
| | OPCOM11 | 0.820 | 6,212 | 0,184 | 1.11 | - |
| | | | | | | |
| **Slovenia** | OPCOM1 | -2.120 | 30.502 | <0.001* | 1.02 | - |
| **CAWI** | OPCOM2 | -4.925 | 7.308 | 0,121 | 0.92 | - |
| | OPCOM3 | -8.006 | 11.917 | 0,018 | 0.83 | - |
| | OPCOM4 | 1.124 | 22.221 | <0.001 | 1.18 | - |
| | OPCOM5 | -3.391 | 20.394 | <0.001 | 1.01 | - |
| | OPCOM6 | -6.280 | 12.622 | 0.013 | 0.89 | agedico, education |
| | OPCOM7 | -2.029 | 6.824 | 0.146 | 1.11 | education*, pay bills* |
| | OPCOM8 | -4.440 | 8.925 | 0.063 | 1 | - |
| | OPCOM9 | -7.518 | 25.793 | <0.001* | 0.81 | - |
| | OPCOM10 | -1.244 | 11.826 | 0.019 | 1.14 | - |
| | OPCOM11 | -1.260 | 3.842 | 0.428 | 1.09 | - |
| | | | | | | |
| **Slovenia[x]** | OPCOM1 | -5.025 | 24.671 | <0.001* | 1.02 | - |
| **CAPI** | OPCOM2 | -7.792 | 2.599 | 0.458 | 0.96 | - |

| Country | Item | Fit.res. | Chi sq *n*=990 | Chi sq *p* | Infit MNSQ | DIF *n*=990 |
|---|---|---|---|---|---|---|
| | OPCOM3 | -11.284 | 15.547 | 0.001 | 0.84 | - |
| | OPCOM4 | -3.548 | 16.233 | 0.001 | 1.16 | - |
| | OPCOM5 | -10.365 | 7.585 | 0.055 | 0.85 | -- |
| | OPCOM6 | -8.980 | 6.403 | 0.094 | 0.89 | - |
| | OPCOM7 | -5.481 | 1.65 | 0.648 | 1.03 | - |
| | OPCOM8 | -9.367 | 15.64 | 0.002 | 0.88 | - |
| | OPCOM9 | -4.006 | 45.754 | <0.001* | 1.08 | - |
| | OPCOM10 | -4.245 | 3.591 | 0.309 | 1.13 | agecat2* |
| | OPCOM11 | -4.397 | 2.019 | 0.568 | 1.07 | - |

[#] non-uniform DIF

*significant when sample size = 660

[x] *df*=3

Not reported in table: Response dependency was observed between items OPCOM1 and OPCOM3 (r= .35) in the German data.

**Table C2b.** $HLS_{19}$-COM-Q6

| **Country** | **Item** | **Fit.res.** | **Chi sq *n*=540** | **Chi sq *p*** | **Infit MNSQ** | **DIF *n*=540** |
|---|---|---|---|---|---|---|
| **Austria** | OPCOM3 | -0.903 | 8.088 | 0.088 | 1.03 | - |
| | OPCOM4 | 3.215 | 12.101 | 0.017 | 1.21 | - |
| | OPCOM5 | -7.609 | 9.920 | 0.042 | 0.88 | - |
| | OPCOM8 | -5.605 | 9.087 | 0.059 | 0.92 | - |
| | OPCOM9 | -3.930 | 3.519 | 0.475 | 0.93 | - |
| | OPCOM10 | 2.536 | 7.092 | 0.131 | 1.15 | - |
| | | | | | | |
| **Austria** | OPCOM3 | -0.375 | 8.893 | 0.064 | 1.06 | - |
| **CAWI** | OPCOM4 | -0.414 | 7.847 | 0.097 | 1.05 | - |
| | OPCOM5 | -5.198 | 10.867 | 0.028 | 0.85 | - |
| | OPCOM8 | -5.017 | 12.118 | 0.017 | 0.88 | - |
| | OPCOM9 | -3.623 | 10.582 | 0.032 | 0.88 | - |
| | OPCOM10 | 5.652 | 29.873 | <0.001 | 1.32 | - |
| | | | | | | |
| **Belgium**[i] | OPCOM3 | 0.276 | 5.107 | 0.277 | 1.07 | - |
| | OPCOM4 | 0.903 | 8.181 | 0.085 | 1.16 | - |
| | OPCOM5 | -4.997 | 16.462 | 0.002 | 0.85 | - |
| | OPCOM8 | -7.954 | 28.393 | <0.001* | 0.76 | - |
| | OPCOM9 | -3.357 | 8.083 | 0.089 | 0.91 | - |
| | OPCOM10 | 3.020 | 19.731 | <0.001 | 1.29 | - |
| | | | | | | |
| **Bulgaria** | OPCOM3 | -2.688 | 10.653 | 0.031 | 0.97 | - |
| CAPI, n=402[a] | OPCOM4 | 3.016 | 23.911 | <0.001* | 1.36 | agecat1*, education*, employment*, soc.level, health |
| | OPCOM5 | -4.676 | 4.155 | 0.385 | 0.80 | - |

| Country | Item | Fit.res. | Chi sq *n*=540 | Chi sq *p* | Infit MNSQ | DIF *n*=540 |
|---|---|---|---|---|---|---|
| | OPCOM8 | -4.983 | 13.011 | 0.011 | 0.79 | agecat1[#] |
| | OPCOM9 | -4.157 | 5.329 | 0.255 | 0.88 | - |
| | OPCOM10 | 0.078 | 5.844 | 0.211 | 1.13 | - |
| | | | | | | |
| **Bulgaria** | OPCOM3 | 0.111 | 25.797 | <0.001* | 1.14 | agecat1*[#], |
| CAWI, n=457 | OPCOM4 | 2.008 | 29.962 | <0.001* | 1.27 | - |
| | OPCOM5 | -4.804 | 9.692 | 0.046 | 0.77 | - |
| | OPCOM8 | -4.853 | 11.482 | 0.022 | 0.82 | - |
| | OPCOM9 | -2.440 | 7.274 | 0.122 | 0.89 | - |
| | OPCOM10 | 0.832 | 8.968 | 0.062 | 1.15 | - |
| | | | | | | |
| **Czechia** | OPCOM3 | -2.250 | 15.824 | 0.003 | 0.97 | - |
| **CAWI** | OPCOM4 | 0.459 | 7.813 | 0.099 | 1.09 | - |
| | OPCOM5 | -6.171 | 10.307 | 0.036 | 0.85 | - |
| | OPCOM8 | -5.683 | 11.631 | 0.020 | 0.86 | - |
| | OPCOM9 | -3.571 | 11.664 | 0.020 | 0.91 | - |
| | OPCOM10 | 3.482 | 19.643 | 0.001 | 1.34 | - |
| | | | | | | |
| **Denmark** | OPCOM3 | -6.843 | 16.221 | 0.002 | 0.95 | - |
| | OPCOM4 | 7.212 | 42.432 | <0.001* | 1.36 | - |
| | OPCOM5 | -15.154 | 18.448 | 0.001 | 0.79 | - |
| | OPCOM8 | -16.249 | 18.971 | 0.001 | 0.77 | - |
| | OPCOM9 | -11.622 | 11.426 | 0.022 | 0.85 | - |
| | OPCOM10 | 4.871 | 22.522 | <0.001* | 1.37 | - |
| | | | | | | |
| **France**[i] | OPCOM3 | -3.915 | 16.762 | 0.002 | 1.02 | - |
| | OPCOM4 | -0.147 | 6.531 | 0.162 | 1.16 | - |
| | OPCOM5 | -4.252 | 10.354 | 0.035 | 0.98 | - |
| | OPCOM8 | -7.4 | 16.791 | 0.002 | 0.9 | - |
| | OPCOM9 | -7.955 | 9.113 | 0.058 | 0.85 | - |
| | OPCOM10 | 0.572 | 7.219 | 0.125 | 1.21 | - |
| | | | | | | |
| **Germany** | OPCOM3 | -1.943 | 16.259 | 0.003 | 1 | - |
| | OPCOM4 | 2.116 | 9.979 | 0.041 | 1.1 | - |
| | OPCOM5 | -5.113 | 4.385 | 0.357 | 0.92 | - |
| | OPCOM8 | -6.135 | 8.21 | 0.084 | 0.88 | - |
| | OPCOM9 | -2.241 | 2.808 | 0.591 | 0.94 | - |
| | OPCOM10 | 3.745 | 10.241 | 0.037 | 1.17 | - |
| | | | | | | |
| **Hungary** | OPCOM3 | -4.453 | 5.619 | 0.230 | 1.01 | - |
| | OPCOM4 | -3.242 | 12.854 | 0.012 | 1.08 | - |
| | OPCOM5 | -7.962 | 11.780 | 0.019 | 0.81 | - |
| | OPCOM8 | -7.449 | 7.902 | 0.095 | 0.86 | - |
| | OPCOM9 | -5.748 | 7.144 | 0.129 | 0.91 | - |

| Country | Item | Fit.res. | Chi sq *n*=540 | Chi sq *p* | Infit MNSQ | DIF *n*=540 |
|---|---|---|---|---|---|---|
| | OPCOM10 | 1.577 | 32.830 | <0.001* | 1.33 | - |
| | | | | | | |
| **Slovenia** | OPCOM3 | -6.303 | 7.088 | 0.131 | 0.88 | - |
| **CAWI** | OPCOM4 | 0.122 | 11.34 | 0.022 | 1.19 | - |
| | OPCOM5 | -4.992 | 13.948 | 0.008 | 0.92 | - |
| | OPCOM8 | -4.487 | 8.882 | 0.064 | 0.99 | - |
| | OPCOM9 | -6.192 | 14.208 | 0.007 | 0.87 | - |
| | OPCOM10 | 1.022 | 15.928 | 0.003 | 1.24 | - |
| | | | | | | |
| **Slovenia** | OPCOM3 | -9.109 | 14.951 | 0.005 | 0.91 | - |
| **CAPI** | OPCOM4 | -4.702 | 5.825 | 0.213 | 1.1 | - |
| | OPCOM5 | -10.789 | 5.324 | 0.256 | 0.83 | - |
| | OPCOM8 | -8.637 | 14.947 | 0.005 | 0.89 | - |
| | OPCOM9 | -5.509 | 22.011 | <0.001 | 1 | - |
| | OPCOM10 | -2.679 | 5.617 | 0.230 | 1.19 | - |

[i] complete data

*significant when sample size = 360

[a]sample size available for Rasch analysis (excluding extreme scorers) when sample size is less than 1000

No response dependency and unordered thresholds were observed for $HLS_{19}$-COM-Q6.

**Table D2.** $HLS_{19}$-NAV single item statistics

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| **Austria** | OPNHL1 | 1.438 | 0.699 | 0.951 | 1.04 | - |
| | OPNHL2 | -3.111 | 4.67 | 0.323 | 0.95 | - |
| | OPNHL3 | 3.355 | 7.469 | 0.113 | 1.11 | agecat1, agecat2, employment |
| | OPNHL4 | -2.634 | 2.824 | 0.588 | 0.96 | - |
| | OPNHL5 | -5.544 | 13.339 | 0.010 | 0.89 | - |
| | OPNHL6 | -2.413 | 1.477 | 0.831 | 0.97 | - |
| | OPNHL7 | -2.119 | 2.221 | 0.695 | 0.96 | education |
| | OPNHL8 | -2.535 | 3.229 | 0.520 | 0.96 | - |
| | OPNHL9 | 6.008 | 36.788 | <0.001* | 1.24 | agedico, agecat1*, agecat2*, employment* |
| | OPNHL10 | -5.158 | 10.881 | 0.028 | 0.88 | - |
| | OPNHL11 | -1.749 | 0.883 | 0.927 | 0.98 | - |
| | OPNHL12 | 1.831 | 4.538 | 0.338 | 1.09 | - |
| | | | | | | |
| **Austria** | OPNHL1 | -0.548 | 2.849 | 0.584 | 1.00 | - |
| **CAWI** | OPNHL2 | -2.692 | 8.952 | 0.062 | 0.93 | - |
| | OPNHL3 | 0.809 | 0.735 | 0.947 | 1.06 | agedico* |
| | OPNHL4 | -0.690 | 5.128 | 0.274 | 0.98 | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| | OPNHL5 | -4.186 | 16.296 | 0.003 | 0.86 | - |
| | OPNHL6 | -2.180 | 2.560 | 0.634 | 0.96 | - |
| | OPNHL7 | -0.711 | 3.730 | 0.444 | 1.01 | - |
| | OPNHL8 | -0.510 | 2.037 | 0.729 | 1.01 | agedico |
| | OPNHL9 | 4.381 | 45.925 | <0.001* | 1.30 | gender*, agedico |
| | OPNHL10 | -4.549 | 20.913 | <0.001 | 0.83 | - |
| | OPNHL11 | 0.059 | 3.503 | 0.478 | 1.03 | - |
| | OPNHL12 | 3.187 | 10.335 | 0.035 | 1.17 | - |
| | | | | | | |
| **Belgium** | OPNHL1 | -0.814 | 7.042 | 0.134 | 0.99 | - |
| | OPNHL2 | -0.585 | 8.495 | 0.075 | 1.01 | - |
| | OPNHL3 | 2.195 | 10.005 | 0.040 | 1.15 | agedico, agecat1, agecat2, employment* |
| | OPNHL4 | -3.166 | 11.127 | 0.025 | 0.91 | - |
| | OPNHL5 | -4.285 | 13.475 | 0.009 | 0.87 | - |
| | OPNHL6 | -0.973 | 2.585 | 0.630 | 1.03 | - |
| | OPNHL7 | -2.007 | 4.300 | 0.367 | 0.96 | agedico, agecat2 |
| | OPNHL8 | -2.749 | 6.961 | 0.138 | 0.89 | agedico, agecat2 |
| | OPNHL9 | 4.387 | 64.875 | <0.001* | 1.39 | - |
| | OPNHL10 | -3.141 | 8.836 | 0.065 | 0.88 | - |
| | OPNHL11 | -0.946 | 2.247 | 0.690 | 0.99 | - |
| | OPNHL12 | 0.662 | 5.624 | 0.229 | 1.08 | pay bills* |
| | | | | | | |
| **Czechia** | OPNHL1 | -1.962 | 3.883 | 0.422 | 0.99 | - |
| **CAWI** | OPNHL2 | -2.653 | 2.387 | 0.665 | 0.94 | - |
| | OPNHL3 | -0.206 | 3.580 | 0.466 | 1.07 | - |
| | OPNHL4 | -1.33 | 7.314 | 0.120 | 1.00 | - |
| | OPNHL5 | -4.23 | 8.569 | 0.073 | 0.90 | - |
| | OPNHL6 | -1.529 | 1.905 | 0.753 | 1.00 | health# |
| | OPNHL7 | -0.837 | 4.046 | 0.400 | 1.05 | gender*, agedico*, agecat2 |
| | OPNHL8 | -3.163 | 4.620 | 0.329 | 0.92 | - |
| | OPNHL9 | -2.817 | 9.393 | 0.052 | 0.92 | - |
| | OPNHL10 | -4.711 | 14.941 | 0.005 | 0.86 | - |
| | OPNHL11 | -2.462 | 2.947 | 0.567 | 0.96 | - |
| | OPNHL12 | 8.143 | 40.942 | <0.001* | 1.35 | agedico, agecat2 |
| | | | | | | |
| **France** | OPNHL1 | -0.449 | 5.691 | 0.223 | 1.05 | - |
| | OPNHL2 | 2.95 | 30.163 | <0.001* | 1.21 | agedico, agecat, agecat2 |
| | OPNHL3 | -0.613 | 4.514 | 0.341 | 1.07 | agedico*, agecat1 agecat2*, employment |
| | OPNHL4 | -3.251 | 14.361 | 0.006 | 0.95 | - |
| | OPNHL5 | -5.057 | 5.574 | 0.233 | 0.91 | - |
| | OPNHL6 | -6.192 | 7.609 | 0.107 | 0.89 | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| | OPNHL7 | -4.327 | 2.390 | 0.664 | 0.95 | agedico*, agecat1*, agecat2* |
| | OPNHL8 | -6.744 | 16.954 | 0.002 | 0.84 | agedico*, agecat2* |
| | OPNHL9 | 8.595 | 133.836 | <0.001* | 1.54 | gender |
| | OPNHL10 | -6.894 | 13.873 | 0.008 | 0.84 | - |
| | OPNHL11 | -4.236 | 3.309 | 0.508 | 0.94 | - |
| | OPNHL12 | -2.100 | 9.500 | 0.050 | 1.00 | - |
| | | | | | | |
| **Germany** | OPNHL1 | -0.817 | 4.704 | 0.319 | 0.97 | - |
| | OPNHL2 | -1.025 | 4.597 | 0.331 | 0.96 | - |
| | OPNHL3 | 3.396 | 12.258 | 0.016 | 1.13 | - |
| | OPNHL4 | -2.877 | 12.326 | 0.015 | 0.91 | - |
| | OPNHL5 | -4.17 | 10.622 | 0.031 | 0.89 | - |
| | OPNHL6 | 2.18 | 16.195 | 0.003 | 1.10 | education |
| | OPNHL7 | -2.201 | 3.496 | 0.479 | 0.98 | - |
| | OPNHL8 | -0.684 | 2.859 | 0.582 | 1.02 | - |
| | OPNHL9 | 1.876 | 9.821 | 0.044 | 1.10 | - |
| | OPNHL10 | -5.575 | 22.911 | <0.001* | 0.85 | - |
| | OPNHL11 | -0.823 | 0.303 | 0.990 | 1.00 | - |
| | OPNHL12 | 3.099 | 10.659 | 0.031 | 1.13 | - |
| | | | | | | |
| **Portugal** | OPNHL1 | -3.567 | 0.31 | 0.989 | 1.05 | - |
| | OPNHL2 | -3.32 | 21.241 | <0.001 | 1.02 | education |
| | OPNHL3 | 1.678 | 15.457 | 0.004 | 1.2 | - |
| | OPNHL4 | -2.695 | 28.25 | <0.001* | 1.05 | agecat1[#], pay bills, employment[#] |
| | OPNHL5 | -4.84 | 2.031 | 0.730 | 0.96 | - |
| | OPNHL6 | -5.247 | 11.959 | 0.018 | 0.96 | pay bills* |
| | OPNHL7 | -6.114 | 20.112 | 0.001 | 0.86 | agedico, agecat1 |
| | OPNHL8 | -5.982 | 28.813 | <0.001* | 0.84 | - |
| | OPNHL9 | -3.65 | 33.106 | <0.001* | 1.09 | pay bills |
| | OPNHL10 | -3.41 | 9.209 | 0.056 | 0.97 | - |
| | OPNHL11 | -4.267 | 10.961 | 0.027 | 0.99 | - |
| | OPNHL12 | -3.443 | 2.828 | 0.587 | 1.03 | - |
| | | | | | | |
| **Slovenia** | OPNHL1 | -1.746 | 2.318 | 0.677 | 1.05 | - |
| **CAPI** | OPNHL2 | -3.597 | 10.094 | 0.039 | 0.95 | - |
| | OPNHL3 | -1.761 | 4.526 | 0.340 | 1.03 | - |
| | OPNHL4 | 2.183 | 33.614 | <0.001* | 1.17 | - |
| | OPNHL5 | -7.081 | 11.787 | 0.019 | 0.87 | - |
| | OPNHL6 | -3.536 | 2.433 | 0.657 | 1 | - |
| | OPNHL7 | -6.574 | 7.148 | 0.128 | 0.87 | - |
| | OPNHL8 | -3.315 | 15.39 | 0.004 | 0.96 | - |
| | OPNHL9 | -3.671 | 28.511 | <0.001* | 1.02 | - |
| | OPNHL10 | -8.211 | 11.998 | 0.017 | 0.82 | - |

| Country | Item | Fit.res. | Chi sq *n*=1080 | Chi sq *p* | Infit MNSQ | DIF *n*=1080 |
|---|---|---|---|---|---|---|
| | OPNHL11 | -3.332 | 3.028 | 0.553 | 1.01 | - |
| | OPNHL12 | 2.316 | 26.632 | <0.001* | 1.21 | - |
| | | | | | | |
| **Slovenia** | OPNHL1 | -1.66 | 17.662 | 0.001 | 1 | - |
| **CAWI** | OPNHL2 | -2.834 | 14.727 | 0.005 | 0.96 | - |
| | OPNHL3 | -2.553 | 1.626 | 0.804 | 0.96 | - |
| | OPNHL4 | -3.101 | 5.204 | 0.267 | 0.95 | - |
| | OPNHL5 | -5.688 | 10.606 | 0.031 | 0.86 | - |
| | OPNHL6 | -3.364 | 2.433 | 0.657 | 0.95 | - |
| | OPNHL7 | -4.145 | 8.528 | 0.074 | 0.93 | - |
| | OPNHL8 | -2.724 | 21.827 | <0.001 | 0.94 | - |
| | OPNHL9 | -1.502 | 32.352 | <0.001* | 1.07 | - |
| | OPNHL10 | -5.472 | 14.242 | 0.007 | 0.87 | - |
| | OPNHL11 | 0.332 | 2.027 | 0.731 | 1.1 | - |
| | OPNHL12 | 9.338 | 75.017 | <0.001* | 1.44 | gender |
| | | | | | | |
| **Switzerland** | OPNHL1 | -1.502 | 3.269 | 0.514 | 1.01 | - |
| **CAWI** | OPNHL2 | -1.534 | 10.058 | 0.040 | 1.00 | - |
| | OPNHL3 | 1.420 | 6.818 | 0.146 | 1.11 | agedico*, agecat1*, agecat2*, employment* |
| | OPNHL4 | -4.599 | 8.145 | 0.086 | 0.92 | - |
| | OPNHL5 | -6.989 | 17.925 | 0.001 | 0.86 | - |
| | OPNHL6 | -4.693 | 2.181 | 0.702 | 0.94 | - |
| | OPNHL7 | -0.441 | 3.028 | 0.553 | 1.06 | education*, pay bills* |
| | OPNHL8 | -2.371 | 3.511 | 0.476 | 0.98 | agedico, education, pay bills |
| | OPNHL9 | 2.711 | 42.022 | <0.001* | 1.22 | gender, agedico, agecat2 |
| | OPNHL10 | -6.54 | 11.014 | 0.026 | 0.86 | - |
| | OPNHL11 | -3.001 | 1.021 | 0.907 | 0.96 | - |
| | OPNHL12 | 0.180 | 2.453 | 0.653 | 1.07 | - |

[#] non-uniform DIF

*significant when sample size = 720

Not reported in table: Response dependency was observed between items OPNHL7 and OPNHL8 in the Belgian (r = .37), Portuguese (r = .43) and Swiss (r = .38) data.

**Table E2.** $HLS_{19}$-VAC single item statistics

| Country | Item | Fit.res. | Chi sq *n=360* | Chi sq *p* | Infit MNSQ | DIF *n=1000*[a] |
|---|---|---|---|---|---|---|
| **Austria** | 19 | -3.105 | 3.888 | .421 | 0.95 | - |
| **CATI** | 22 | -6.739 | 9.335 | .053 | 0.87 | - |
| | 26 | -6.364 | 6.742 | .150 | 0.88 | - |
| | 29 | 8.827 | 28.200 | <.001* | 1.33 | - |
| | | | | | | |
| **Austria** | 19 | -3.411 | 1.745 | .783 | 0.71 | - |

| Country | Item | Fit.res. | Chi sq *n=360* | Chi sq *p* | Infit MNSQ | DIF *n=1000*[a] |
|---|---|---|---|---|---|---|
| **CAWI** | 22 | -7.286 | 6.481 | .166 | 0.60 | - |
| | 26 | -7.890 | 5.690 | .224 | 0.52 | - |
| | 29 | 1.831 | 12.582 | .014 | 0.75 | - |
| | | | | | | |
| **Belgium** | 19 | -1.280 | 6.747 | .150 | 0.98 | agedico[a] (non-unif.) |
| n=868 | 22 | -4.893 | 3.585 | .465 | 0.89 | - |
| | 26 | -4.659 | 8.598 | .072 | 0.84 | - |
| | 29 | 4.161 | 12.944 | .012 | 1.37 | agecat1[a] |
| | | | | | | |
| **Bulgaria** | 19 | 2.349 | 5.663 | .226 | 1.17 | - |
| **CAPI** | 22 | -1.424 | 4.720 | .317 | 0.91 | agecat1[a] |
| n=379 | 26 | -2.137 | 1.924 | .027 | 0.85 | edudico[a] |
| | 29 | .739 | 4.754 | .313 | 1.06 | - |
| | | | | | | |
| **Bulgaria** | 19 | 5.261 | 43.893 | <.001* | 1.26 | agecat1 (non-unif.)[a], health[a] |
| **CAWI** | 22 | -2.418 | 18.582 | .001 | 0.88 | - |
| n=442 | 26 | -2.200 | 17.234 | .002 | 0.88 | employment[a] |
| | 29 | .099 | 5.360 | .252 | 1.01 | - |
| | | | | | | |
| **Czechia** | 19 | -1.497 | 1.763 | .623 | 0.96 | agecat1[a] |
| **CAWI** | 22 | -1.276 | 1.337 | .720 | 0.98 | - |
| n=977 | 26 | -4.166 | 9.856 | .020 | 0.84 | - |
| | 29 | 3.055 | 5.306 | .151 | 1.23 | employment, billsdico, soclevel |
| | | | | | | |
| **Czechia** | 19 | -1.911 | 2.325 | .676 | 1.01 | agecat1[a], agedico[a], agecat2[a], employment[a] |
| **CATI** | 22 | -3.621 | 13.105 | .011 | 0.81 | agecat1[a], billsdico[a] |
| n=471 | 26 | -2.386 | 1.083 | .039 | 0.89 | - |
| | 29 | 2.142 | 14.400 | .006 | 1.29 | agedico[a], agecat2[a], edudico[a], billsdico[a], health[a] |
| | | | | | | |
| **Germany** | 19 | .120 | .707 | .951 | 0.99 | edudico |
| | 22 | -4.164 | 4.932 | .294 | 0.90 | agedico, agecat1, agecat2 |
| | 26 | -4.398 | 1.894 | .028 | 0.88 | gender |
| | 29 | 7.644 | 8.298 | .081 | 1.25 | agedico, agecat1, agecat2, edudico, employment |
| | | | | | | |
| **Hungary** | 19 | -2.043 | 3.683 | .298 | 1.07 | soclevel (non-unif.), health |
| | 22 | -5.346 | 3.264 | .353 | 0.89 | - |
| | 26 | -5.534 | 6.199 | .102 | 0.91 | edudico[a] (non-unif.) |
| | 29 | -.136 | 3.859 | .277 | 1.20 | soclevel (non-unif.)[a], billsdico |

| Country | Item | Fit.res. | Chi sq *n=360* | Chi sq *p* | Infit MNSQ | | DIF *n=1000*[a] |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Ireland** | | | | | Set A | Set B | |
| N > 3000 | 19 | 1.931 | 1.865 | .761 | 0.99 | 1.00 | employment |
| | 22 | -2.382 | 8.155 | .086 | 0.97 | 0.96 | - |
| | 26 | -1.494 | 8.910 | .063 | 0.95 | 0.93 | gender |
| | 29 | 6.290 | 3.953 | .412 | 1.20 | 1.20 | agedico, agecat1, agecat2, employment |
| | | | | | | | |
| **Italy** | 19 | -2.899 | 9.465 | .050 | 0.98 | | billsdico(non-unif.)[a] |
| **CATI** | 22 | -3.351 | 1.893 | .755 | 0.99 | | - |
| n=529 | 26 | -2.176 | 5.682 | .128 | 0.98 | | agecat2[a], employment[a], soclevel (non-unif.)[a] |
| | 29 | -1.821 | 6.858 | .077 | 1.10 | | - |
| | | | | | | | |
| **Italy** | 19 | .271 | 2.503 | .475 | 1.03 | | - |
| **CAWI** | 22 | -4.586 | 5.136 | .162 | 0.95 | | agecat1 |
| | 26 | -2.509 | 1.764 | .623 | 0.97 | | Gender |
| | 29 | .502 | 1.282 | .733 | 1.06 | | agecat1 |
| | | | | | | | |
| **Norway** | 19 | 1.671 | 3.087 | .543 | 1.03 | | - |
| | 22 | -3.580 | 13.936 | .008 | 0.97 | | agedico[a], agecat1[a], agecat2[a], employment[a] |
| | 26 | -4.694 | 12.754 | .013 | 0.88 | | gender (non-unif.), agecat1, edudico |
| | 29 | 6.925 | 6.376 | .173 | 1.18 | | gender (non-unif.), agedico[a], agecat1[a], agecat2[a], employment[a] |
| | | | | | | | |
| **Portugal** | 19 | -5.724 | 12.816 | .005 | 0.94 | | agedico, agecat1, agecat2, employment[a] |
| | 22 | -7.068 | 3.601 | .463 | 0.88 | | agedico[a], agecat1[a] (non-unif.), agecat2[a], edudico, employment[a], soclevel[a] |
| | 26 | -7.133 | 31.660 | <.001* | 0.90 | | - |
| | 29 | -.514 | 26.117 | <.001* | 1.21 | | agedico[a], agecat1[a], agecat2[a], edudico, employment[a], billsdico, soclevel[a], health |
| | | | | | | | |
| **Slovenia** | 19 | 1.580 | 2.592 | .628 | 1.05 | | - |
| **CAWI** | 22 | -2.261 | 8.115 | .087 | 0.93 | | billsdico |
| | 26 | -3.211 | 11.500 | .021 | 0.85 | | - |
| | 29 | 4.664 | 7.103 | .131 | 1.20 | | health |
| | | | | | | | |
| **Slovenia** | 19 | -3.670 | 1.600 | .659 | 0.96 | | agecat1, agecat2[a], employment |
| **CAPI** | 22 | -5.148 | 6.439 | .092 | 0.92 | | edudico |

| Country | Item | Fit.res. | Chi sq *n=360* | Chi sq *p* | Infit MNSQ | DIF *n=1000*[a] |
|---|---|---|---|---|---|---|
| | 26 | -6.328 | 3.158 | .368 | 0.89 | - |
| | 29 | 2.529 | 9.493 | .023 | 1.24 | agecat1, agecat2, edudico, employment |

*significant when sample size = 240, [a]owing to few items we used amend sample size = 1000 for DIF-analysis (or actual sample size when less than 1000), [a]DIF significant even for sample size = 500