



HAL
open science

The Stata module -pcm-for Rasch Measurement Theory -An application to assess Health Literacy determinants for residents of homeless shelters

Jean-Benoit Hardouin, Yseulys Dubuy, Myriam Blanchin, Simon Ducarroz

► To cite this version:

Jean-Benoit Hardouin, Yseulys Dubuy, Myriam Blanchin, Simon Ducarroz. The Stata module -pcm-for Rasch Measurement Theory -An application to assess Health Literacy determinants for residents of homeless shelters. 2024. hal-04812388

HAL Id: hal-04812388

<https://hal.science/hal-04812388v1>

Preprint submitted on 30 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Stata module -pcm- for Rasch Measurement Theory - An application to assess Health Literacy determinants for residents of homeless shelters

Jean-Benoit Hardouin
INSERM UMR 1136-IPLeSP and UMR 1246-SPHERE
Nantes University
Nantes/France

jean-benoit.hardouin@univ-nantes.fr

Yseulys Dubuy	Myriam Blanchin
INSERM UMR 1246-SPHERE	INSERM UMR 1246-SPHERE
Nantes University	Nantes University
Nantes/France	Nantes/France
yseulys.Dubuy@univ-nantes.fr	myriam.blanchin@univ-nantes.fr

Simon Ducarroz
INSERM UMR 1136-IPLeSP
Sorbonne University
Paris/France
simon.ducarroz@inserm.fr

Abstract. The document presents the Stata command `pcm` designed for Partial Credit Models (PCM) within Rasch Measurement Theory (RMT), aimed at analyzing questionnaire data with polytomous items. The paper details the `pcm` command's use in assessing health literacy determinants among homeless shelter residents, particularly examining the "Social Support for Health" dimension from the Health Literacy Questionnaire (HLQ).

The `pcm` command addresses RMT limitations in existing Stata tools, offering quick parameter estimation, automatic generation of fit indices, and graphical outputs to simplify analysis. Key features include the estimation of latent trait levels, fit assessment using INFIT and OUTFIT indices, and visualization through Item and Test Characteristic Curves (ICCs and TCCs). The command also supports latent regression, enabling analysis of external predictors' impact on the latent trait.

An empirical illustration using data from the 2021 ECHO study demonstrates the `pcm` command's capabilities. It highlights the role of health perception, language proficiency, and social factors in health literacy among the homeless, with results indicating good model fit and reliability (PSI = 0.78). Variables such as French language skills, sex, and perceived health status were significant predictors of health literacy levels.

Keywords: `st0001`, `pcm`, Rasch Measurement Theory (RMT), Partial Credit Model (PCM), Rating Scale Model (RSM)

1 Introduction

Rasch's family models comprises a set of latent variable models widely utilized in psychology, educational sciences and health Christensen et al. (2012). These models enable the measurement of latent traits in individuals based on their responses to questionnaires or ability tests composed of several dichotomous or polytomous items. In these models, the latent trait is typically considered as a random variable, and the items are characterized by several parameters, referred to threshold parameters in this paper, although they are often known as difficulty parameters.

Rasch's family models verify two main properties Christensen et al. (2012) Bond and Fox (2007):

- Specific objectivity property: This property derives from the parameter separability in the Rasch models. That means that individual parameters (distribution parameters of the latent characteristics) can be estimated independently from item parameters (corresponding to thresholds for these items) and conversely. Practically, this means the sample's level on the latent trait does not affect the estimation of the threshold parameters, and conversely, the item threshold parameters do not influence the estimation of the individuals' latent trait. As a consequence, the specific objectivity property refers to the invariance of comparisons of individuals or items. Hence, comparisons of individual levels of latent traits (item parameters respectively) is independent from the selected set of items (the set of individuals respectively). This property, for example, enables simple and efficient handling of missing data, even when the missing data are informative Hardouin et al. (2011) Hamel et al. (2017).
- Sufficiency of raw scores to estimate the latent variable parameters: all individuals with the same raw score (defined as the unweighted sum of responses to all items of a given dimension) will have the same estimation of the latent variable, regardless of their specific response patterns. This property simplifies linking between the easily computed raw scores and the latent variable levels, which requires estimation via the model.

Rasch Measurement Theory (RMT) posits that models satisfying these two properties possess optimal psychometric characteristics. Therefore, in this theory, data from a well-constructed questionnaire must fit well with a Rasch model. This perspective contrasts with Item Response Theory (IRT), where the primary objective is to identify the model that best fits the data, even if it does not exhibit these two properties Andrich and Marais (2019). The IRT approach is more conventional among statisticians, as it focuses on finding the best-fitting model, often leading to sophisticated models designed to enhance fit.

The estimation and evaluation of fit for Rasch family models within Stata have been extensively documented in Zheng and Rabe-Hesketh (2007), Hardouin (2007), and Hamel et al. (2016). Since Stata 14, the `irt` command ([IRT] `irt`) has been available to support parameter estimation for various Rasch models. However, these approaches

Command	<code>gllamm</code>	<code>irt</code>	<code>uirt</code>	<code>raschtest</code>	<code>pcmodel</code>	<code>pcm</code>
IRT framework		X	X			
RMT framework	X			X	X	X
Polytomous items	X	X	X		X	X
Parameters estimation	X	X	X	X	X	X
Fit assessment			X	X	X	X
Easy DIF analysis		X	X			X
Graphical representations		X	X	X		X
Quick estimation process		X	X	X		X

Table 1: Comparison of the characteristics of the Stata commands for estimation of IRT and RMT models

reveal several limitations from the perspective of Rasch Measurement Theory (RMT).

For instance, Zheng and Rabe-Hesketh (2007) introduced an IRT-based methodology using the `gllamm` command, which focuses solely on estimating model parameters without offering fit indices/tests or graphical representations. The `irt` command, while providing valuable graphical outputs, falls short in terms of fit indices/tests and includes some unconventional options for RMT practitioners. Specifically, the [IRT] `irt` command requires the estimation of a discrimination parameter, which deviates from traditional RMT practices and affects the interpretation of item parameters. This approach necessitates fixing the variance of the latent variable, leading to a conflation of item characteristics and population heterogeneity regarding the latent variable.

Moreover, the [IRT] `uirt` command introduced by Kondratek (2022) extends the functionality of [IRT] `irt` but maintains a similar methodological framework. Additionally, these commands do not fully exploit the unique properties of Rasch models when presenting results. For example, they fail to provide a table that links raw scores to the latent variable using the sufficiency property of the sum score, a standard practice for establishing this critical connection.

Within the specific RMT framework, the `raschtest` command (Hardouin (2007)) is limited to the analysis of dichotomous items, while the `pcmodel` command (Hamel et al. (2016)) uses `gllamm` command and requires considerable time to run.

The new `pcm` command, based on the estimation process of the `gsem` command ([SEM] `sem`), offers a quick estimation of parameters. Additionally, this command automatically generates several graphs and tables aligned with the RMT framework. Table 1 compares all the Stata commands for RMT and IRT models.

In this paper, we demonstrate the application of the new Stata command `pcm` using data that measures Health Literacy. Health Literacy refers to an individual's abilities and resources to access, understand, and evaluate health-related information and to take action to improve health. Our analysis seeks to identify the determinants of Health Literacy among residents of homeless shelters.

2 The Rasch's family models

The Rasch family includes three models: the Rasch model (designed for dichotomous items only) Rasch (1980), the Rating Scale Model (RSM) Andrich (1978), and the Partial Credit Model (PCM) Masters (1982), with the last two able to handle polytomous items. Among these, the PCM is the most flexible model, with both the Rasch model and the RSM being special cases of the PCM. Therefore, we place a particular emphasis on the PCM in the remainder of this manuscript.

2.1 Estimation of the threshold parameters in Rasch's family models

The Partial Credit Model (PCM)

Consider a set of J items X_1 to X_J , each with $K_j + 1$ ordered response categories (coded from 0 to K_j). Each individual n of the sample of size N provides their response to each item denoted as x_{nj} , $\forall j = \{1, \dots, J\}$. The PCM models the response process to all these items using the probabilistic model defined in (1)

$$P(X_{nj} = k | \theta, \delta_j) = \frac{\exp(k\theta - \sum_{l=1}^k \delta_{jl})}{\sum_{m=1}^{K_j} \exp(m\theta - \sum_{l=1}^m \delta_{jl})} \quad (1)$$

where θ is a random variable representing the latent characteristic of the individuals, measured over the range of the real numbers, and $\delta_j = (\delta_{j1}, \dots, \delta_{jK_j})$ is a set of threshold parameters defining the response process for item j .

The figure 1 represents the Categories Characteristic Curves (CCC), which provide a graphical representation of the probabilistic model described in equation (1). The dashed lines represent the values of threshold parameters for this item having 3 responses categories.

In the `pcm` command, the parameters of the model are estimated using Marginal Maximum Likelihood (MML) Fischer and Molenaar (1995), with the latent variable treated as a random variable. This approach assumes the latent variable follows a normal distribution, applying the standard identifiability constraint of fixing its mean at 0. Consequently, only the variance of the latent variable is estimated. The estimation is carried out using StataCorp's `gsem` command (Generalized Structural Equation Modeling).

The Rating Scale Model (RSM)

The Rating Scale Model (RSM) is a special case of the Partial Credit Model (PCM) designed to simplify parameter estimation when all items in a questionnaire or test have the same number of response categories (and the same answer choices), denoted as $K + 1$ Christensen et al. (2012). In the RSM, the differences between threshold parameters are assumed to be constant across items. Specifically, the difference parameters are defined as:

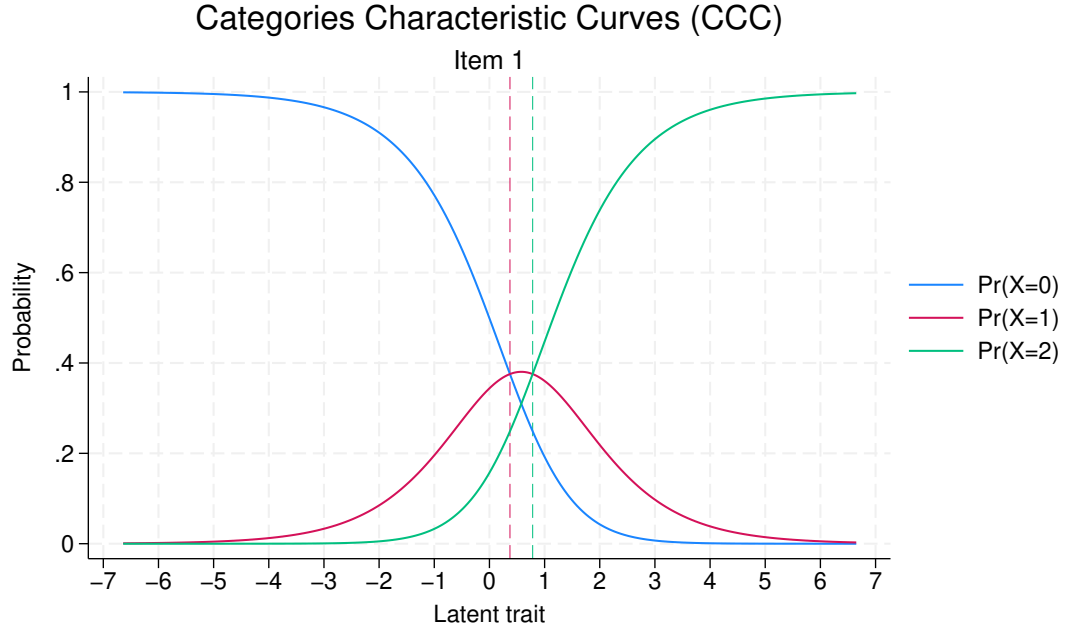


Figure 1: Categories Characteristic Curves (CCC) of an item with 3 response categories (0, 1, 2)

$$\tau_k = \delta_{jk} - \delta_{j1} \quad \forall j = \{1, \dots, J\} \text{ and } k = \{2, \dots, K\} \quad (2)$$

This model is more parsimonious than the PCM because it requires estimating only one specific threshold parameter per item (δ_{j1}) and only $K - 1$ τ_k parameters. Consequently, the total number of parameters to estimate in the RSM is $J + K - 1$, which is always lesser than or equal to the $\sum_{j=1}^J K_j$ parameters required for the PCM.

Case of dysfunctioning items

A dysfunctioning item is one in which the threshold parameters are improperly ordered. This means that the estimated parameters do not adhere to the required sequence:

$$\hat{\delta}_{j1} \leq \hat{\delta}_{j2} \leq \dots \leq \hat{\delta}_{jK_j} \text{ for the PCM} \quad (3)$$

$$0 \leq \hat{\tau}_2 \leq \hat{\tau}_3 \leq \dots \leq \hat{\tau}_K \text{ for the RSM.} \quad (4)$$

When an item is dysfunctioning, one or more of its response categories are not

properly ranked, meaning that a particular response category is never the most likely choice, regardless of the level of the latent variable.

In practice, this implies that such a response category is seldom selected by individuals. This issue is illustrated in Figure 2, which depicts the CCC for a dysfunctional item with four response categories. In this example, response category 1 (shown by the red curve) is never the most probable.

To address this issue, it is often necessary to collapse the response categories of dysfunctional items to reduce their number. For example, responses coded as 1 and 2 might be combined into a new response category.

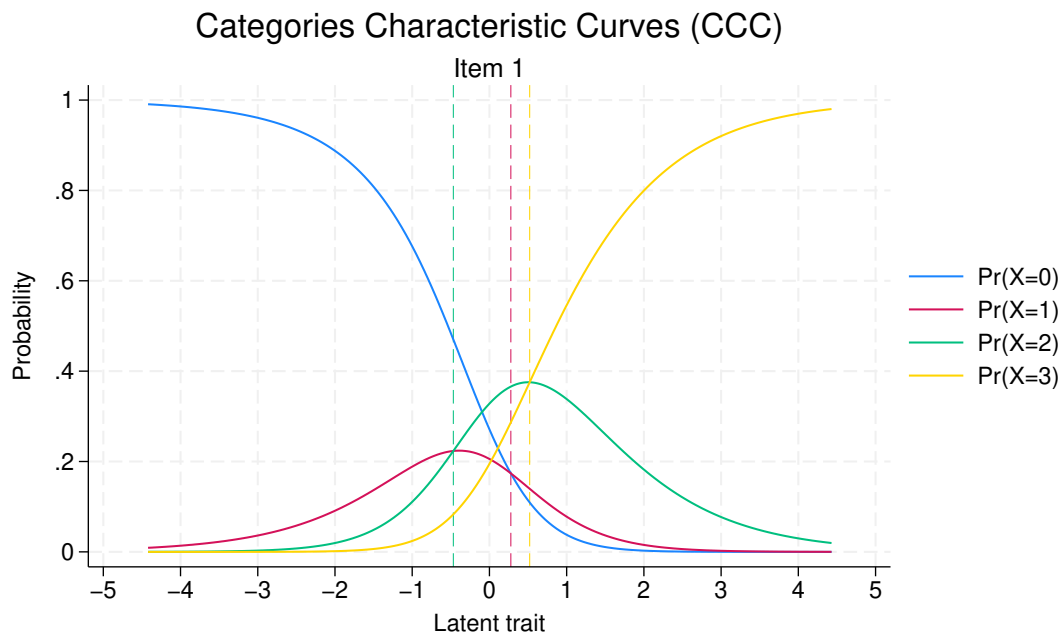


Figure 2: Categories Characteristic Curves (CCC) of a dysfunctional item ($\hat{\delta}_{j1} = 0.28$, $\hat{\delta}_{j2} = -0.47$ and $\hat{\delta}_{j3} = 0.52$ so $\hat{\delta}_{j1} > \hat{\delta}_{j2}$)

2.2 Estimation of the individual values of the latent variable

After estimating the item threshold parameters, several methods can be used to estimate individual latent variable levels. The `pcm` command in Stata offers five distinct types of estimations Fischer and Molenaar (1995):

- Expected A Posteriori (EAP) Estimates: These can be obtained using the `predict` posterior command following `gsem`. This method give the value of the latent

variable that maximizes the probability to obtain a response pattern similar to the one observed for each individual.

- **Plausible Values (PV):** These are derived by drawing a random value from the distribution of the EAP for each individual. Specifically, a random value is sampled from a normal distribution with the EAP as the mean and the standard error of the EAP as the standard deviation.
- **Corrected Estimates (CE):** These estimates determine the latent variable value that yields the same expected raw score as the observed raw score for each individual with a complete response pattern. For individuals with missing data, a linear interpolation between conditional expectation (CE) and expected a posteriori (EAP) estimates is then applied.
- **Most Likely Estimates (MLE):** For each item, MLE calculates the latent variable value that maximizes the probability of the individual's responses to each category. The average of these values is then computed. Note that this method may not yield consistent values for extreme response categories (those coded as 0 or K_j for item j). Therefore, it is recommended to estimate the latent variable for responses around 0.25 or $K_j - 0.25$, as suggested by Winsteps software Wright (1998).
- **Weighted Most Likely Estimates (WMLE):** This method adjusts the MLE by weighting probabilities according to the density function of the latent trait, which is effective across all response categories.

Among these methods, only the EAP and CE approaches fully satisfy the property of score sufficiency for the latent variable, whereas the other methods may yield different latent variable estimates for individuals with identical scores. Notably, the EAP method tends to concentrate estimates around the mean of the latent trait, which can skew model fit evaluation. Therefore, the CE method is used as the default method for the `pcm` command. However, certain graphs can be displayed as a function of EAP, and all these estimates can be saved.

2.3 The assessment of the reliability

In RMT and IRT, reliability is typically assessed using the Person Separation Index (PSI) Wright and Stone (1999) Andrich and Marais (2019). The PSI is calculated as follows:

$$PSI = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\eta}^2} \quad (5)$$

Here, $\hat{\eta}^2$ represents the mean squared value of the standard errors of the individual latent trait estimates, while $\hat{\sigma}^2$ denotes the estimated variance of the latent variable.

Thus, $\hat{\eta}^2$ reflects the measurement error, whereas $\hat{\sigma}^2$ indicates the variability of individuals on the latent trait.

Typically, a PSI value greater than 0.7 is considered to indicate acceptable reliability. However, for individual evaluations with significant personal implications (such as diagnostic assessments or student evaluations), PSI values exceeding 0.8 or 0.9 are often desired.

2.4 Fit assessment

OUTFIT and INFIT indices are used to assess the fit between a measurement model and the observed data Linacre (2022) Müller (2020). These indices derive from residuals.

We define z_{nj} as the standardized residual for the response of individual n to item j , computed as:

$$z_{nj} = \frac{x_{nj} - E(X_{nj})}{\sqrt{V(X_{nj})}} \quad (6)$$

where $E(X_{nj})$ represents the a posteriori expected value of the response from individual n to item j , based on the estimated latent variable for individual n and the threshold parameters for item j , and $V(X_{nj})$ denotes the estimated variance of this expected value.

Figure 3 illustrates the standardized residuals z_{nj} for an item as a function of the latent variable. The color of the points indicates the response category (ranging from 0 to 3). Vertical lines on the graph represent the item threshold parameters, and the number of individuals with large standardized residuals is also displayed. Since the standardized residuals are expected to follow a standard normal distribution, bounds representing the selected percentiles of this distribution are shown on the graph.

$$OUTFIT_j = \frac{1}{N} \sum_{n=1}^N z_{nj}^2 \quad (7)$$

The INFIT index for item jj is calculated as:

$$INFIT_j = \frac{\sum_{n=1}^N V(X_{nj}) z_{nj}^2}{\sum_{n=1}^N V(X_{nj})} \quad (8)$$

For a well-fitting model, both OUTFIT and INFIT indices should be close to 1. Various thresholds can be applied to detect overfitting (high values) or underfitting (low values). For instance, Bond and Fox (2007) suggests that values should fall between 0.75 and 1.30, while Smith et al. (1998) recommends that INFIT values be within $1 \pm 2/\sqrt{N}$ and OUTFIT values within $1 \pm 6/\sqrt{N}$ with N denotes the sample size. Notably, the bounds proposed by Smith et al. (1998) are consistent with findings from a simulation study Müller (2020).

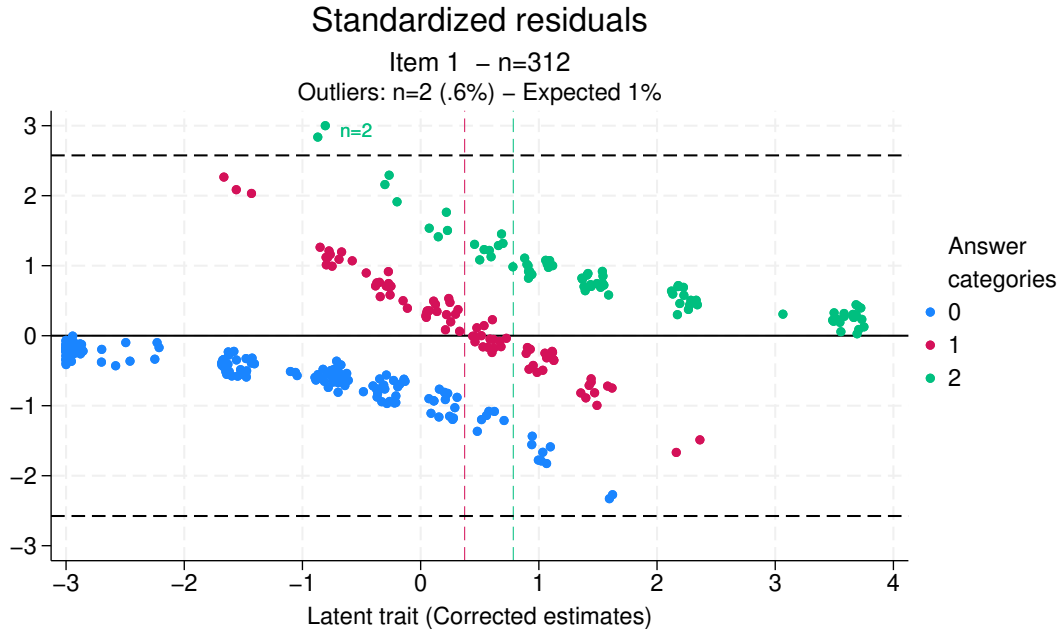


Figure 3: Standardized residuals of an item

Underfitting signifies a poor model fit, whereas overfitting often points to issues with the dimensionality of the item set, which may include multiple sub-dimensions. The OUTFIT index is particularly sensitive to unexpected responses from individuals whose latent variable levels are far from the item's threshold parameters. In contrast, the INFIT index is more responsive to unexpected responses from individuals whose latent variable levels are close to the item's threshold parameters.

Both indices can be standardized to approximate a normal distribution, though this transformation is sometimes debated.

Model fit can also be assessed through graphical representations such as Item Characteristic Curves (ICCs - see Figure 4) or Test Characteristic Curves (TCCs - see Figure 5). These curves illustrate the expected score for each item or for the test (entire questionnaire) as a function of the latent trait. To evaluate the model fit, individuals are grouped based on their estimated latent variable values. For each group, a bullet (with size proportional to the group size) is plotted, showing the average latent trait value on the x-axis and the average observed response (ICC) or score (TCC) on the y-axis. Each bullet is accompanied by a segment that represents the heterogeneity within the group, reflecting variations among individuals with different levels of the latent trait. If the theoretical curve diverges too far from the observed points, the fit may be considered inadequate.

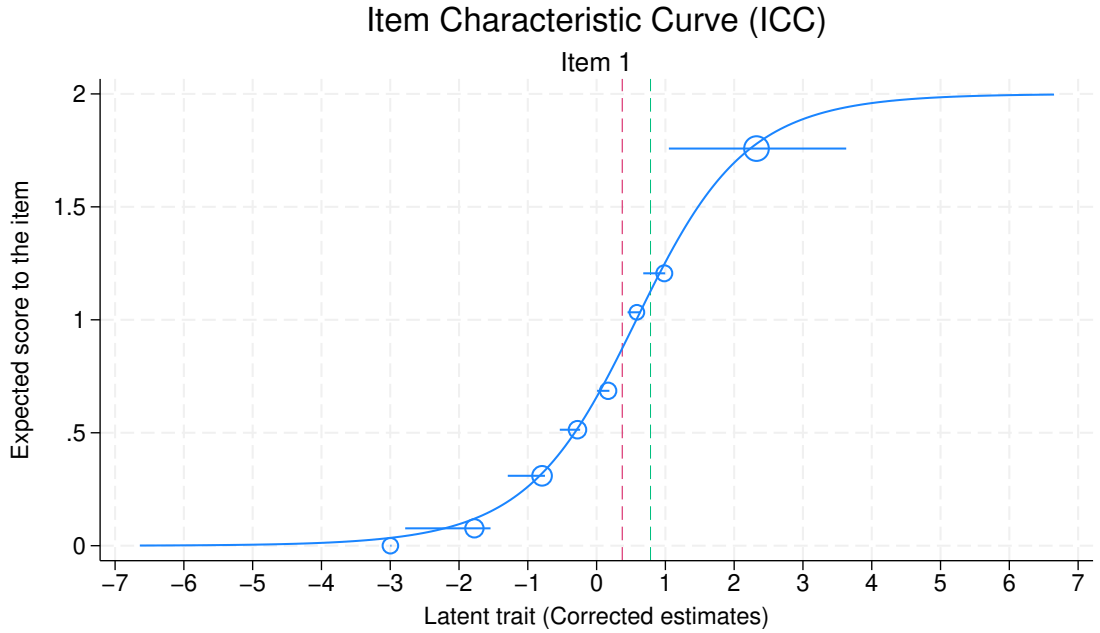


Figure 4: Item Characteristic Curves (ICC) of an item

2.5 Local independence assessment

The Rasch model relies on the assumption of local independence among items, positing that item responses are primarily influenced by the latent variable, which requires that residuals for each item be independent. This assumption is commonly evaluated through principal component analysis (PCA) on standardized residuals. Local independence is considered satisfied if no substantial eigenvalues are observed in the PCA. A parallel analysis can be performed for identifying a large value: a set of simulated datasets with independent residuals (standard normal variables equal in number to the studied items) undergoes PCA, and the k-th percentile (typically the 95th) of the distribution of the first eigenvalues is established as the threshold for identifying a large value.

2.6 The information curves

The Item Information Curves (IIC) 6 are derived from the second derivative of the Categories Characteristic Curve (CCC) Fischer and Molenaar (1995). Higher values of the IIC indicate greater reliability in estimating the latent variable. The Test Information Curve (TIC - see Figure 7) represents the aggregate of all items information curves. The IIC/TIC can help identify redundant items: if multiple items peak at the same latent trait levels, they provide overlapping information, which may be considered redundant.

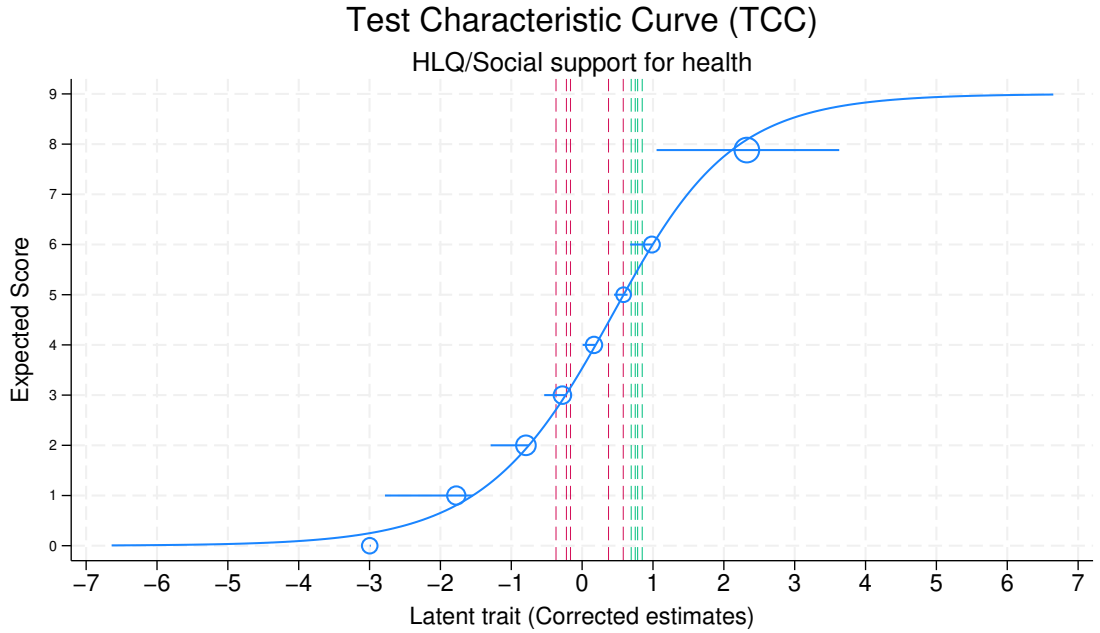


Figure 5: Test Characteristic Curves (TCC)

The Items Information Ratios (IIR - see Figure 8) reflect the rates of information each item contributes at different levels of the latent trait.

2.7 The adapted Wright map

The Wright map Bond and Fox (2007) (see Figure 9) is a traditional graphical representation of RMT analysis. In this visualization, a bar chart (at the top of the graph) depicts the distribution of individuals along the latent variable, while points (at the bottom of the graph) represent the threshold parameters of the items. In the `pcm` command, this representation is enhanced by adding the distribution function of the latent variable (purple curve) and the test information curve (blue curve). We also distinguish between individuals with complete response patterns (depicted by light pink bars) and those with incomplete patterns (depicted by red bars). This graph effectively separates information about individuals (at the top) from information about items (at the bottom). A close alignment between the green and red curves indicates that the test fits the sample well—neither too difficult nor too easy—whereas significant discrepancies suggest a poor fit.

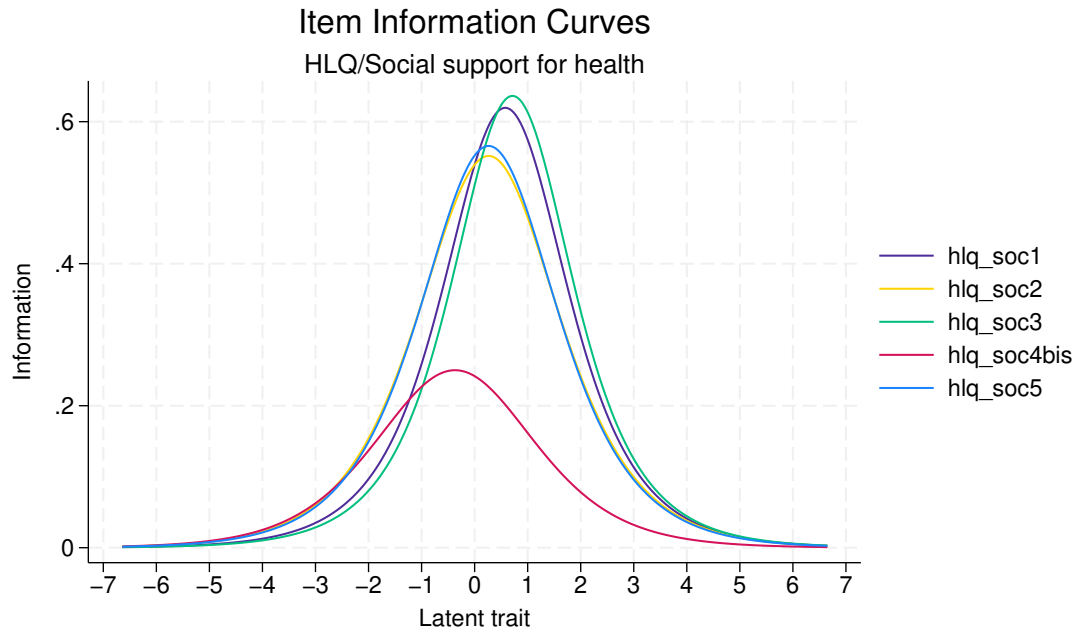


Figure 6: Items information curves (IIC)

2.8 Latent regression to explain the latent variable

Latent regression introduces variables that explain the latent variable. In this approach, the random variable θ in Equation 1 is replaced by a linear model that includes continuous or categorical predictors. The residuals of this model represent the portion of the latent variable that cannot be explained by the included predictors.

3 The `pcm` command

We introduce the new Stata command `pcm`, which automatically generates tables and graphs within the RMT framework.

3.1 Syntax

The syntax of this command is:

```
pcm varlist [if] [in] [, rsm iterate(#) model difficulty(matrixname)
  variance(#) minsize(#) pca pcasim(#) pcacentile(#) graphs
```

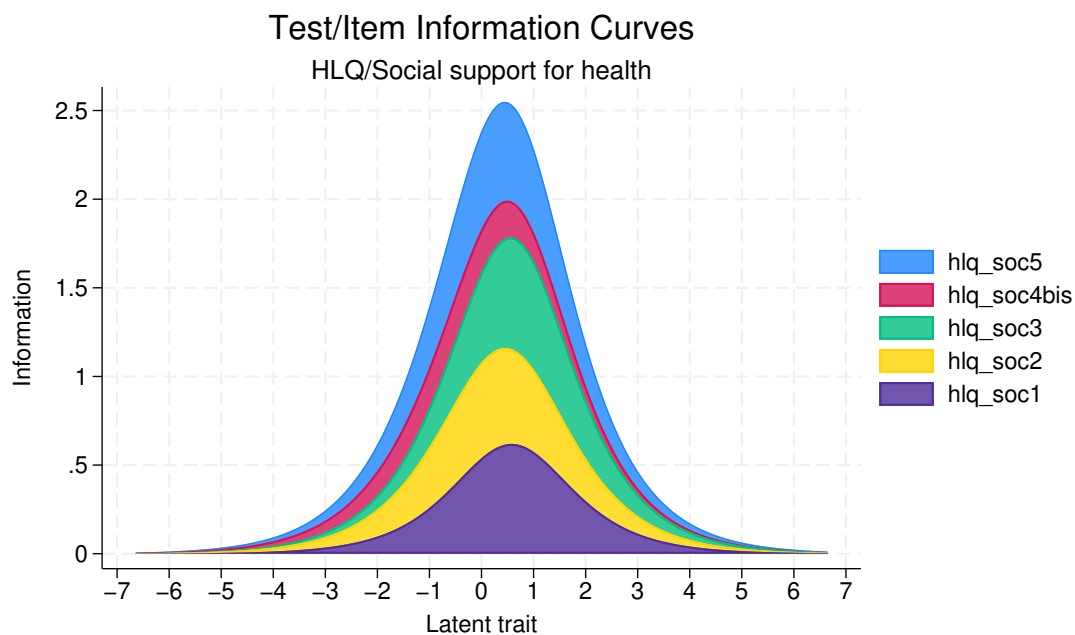


Figure 7: Test information curves (TIC)

```
nographitems dimname(string) noobs nocorrected eap mle noresiduals
jitter(#) alpha(#) wcc filesave extension(string) dirsave(directory)
docx(string) genlt(string) geninf(string) replace ]
```

The user-written commands `genscore` and `gengroup` must be installed beforehand; both commands are available from SSC.

3.2 Options

The following options can be used:

`rsm` estimates a Rating Scale Model instead of a Partial Credit Model.

`iterate` specifies the maximum number of iterations for the maximization algorithm, which defaults to 100.

`model` displays the outputs of the maximization algorithm.

`difficulty` sets the values for the difficulty/threshold parameters of the items (these are estimated by default). A matrix containing threshold parameter values should be defined ahead: this matrix should have a number of rows equal to the number

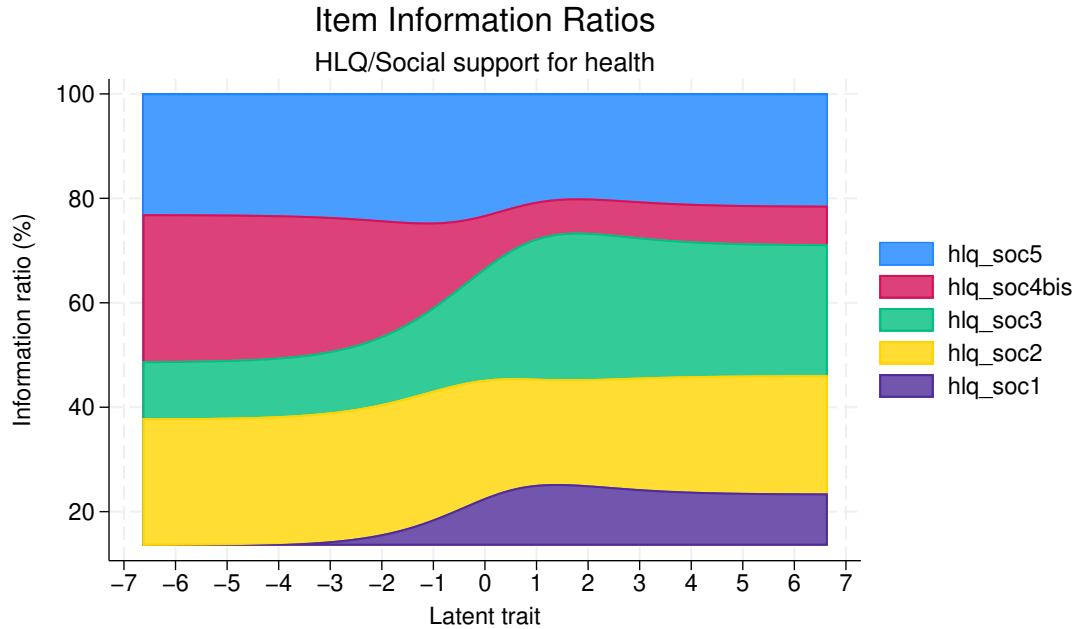


Figure 8: Items information ratios (IIR)

of items, with a column for each threshold parameter (number of response categories minus one). Missing values (.) should be used to indicate non-existent difficulty/threshold parameters in the matrix.

variance fixes the variance of the latent trait to a specified value (default is to estimate this value).

minsize sets the minimum size for the groups (default is 30 individuals). This value may be automatically adjusted to optimize graph outputs.

pca performs a Principal Component Analysis (PCA) on standardized residuals.

pcassim performs a Parallel Analysis in which # sets of simulated datasets, composed of independent standardized normal variables matching the number of residuals and with a number of individuals equal to the sample size, are analyzed to extract the first eigenvalues of the PCA. A specific percentile of the distribution of the first eigenvalues is displayed. By default, no simulation is performed.

pcacentile specifies the percentile of the distribution of the first eigenvalues in the parallel analysis to display. By default, the 95th percentile is shown.

continuous specifies a list of continuous variables that explain the latent trait.

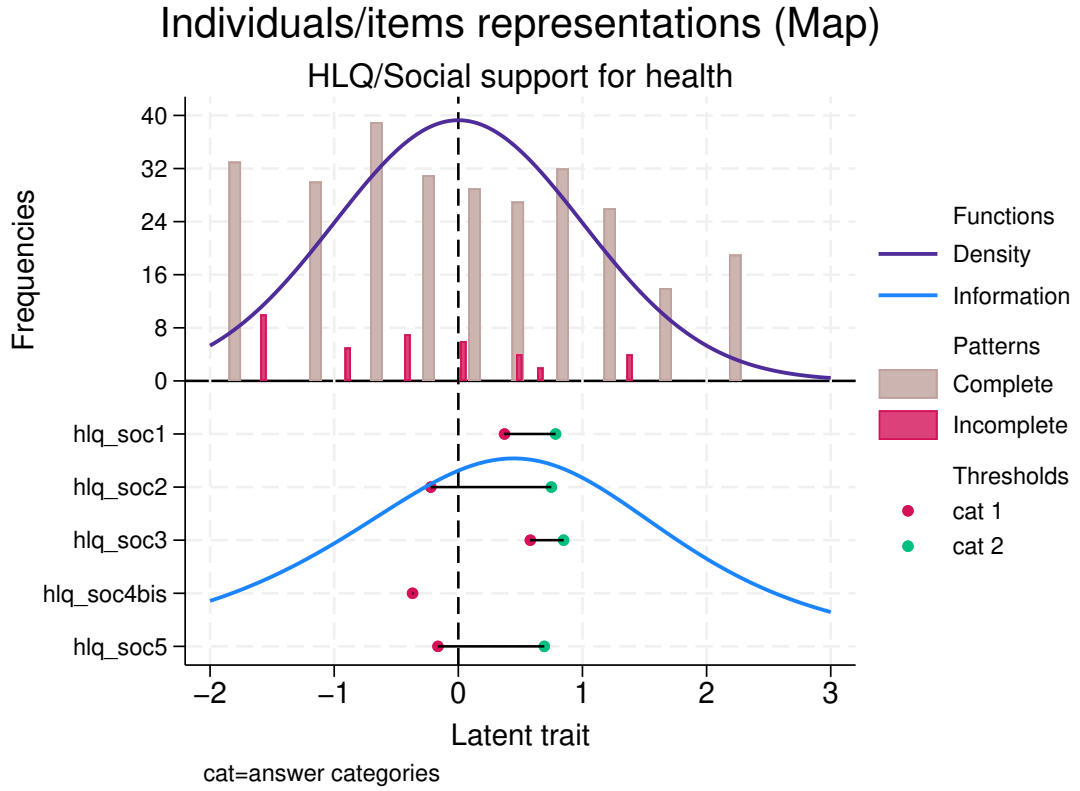


Figure 9: Adapted Wright map

`categorical` specifies a list of categorical variables that explain the latent trait.

`graphs` displays all the graphs

`nographitems` suppresses the display of item-specific graphs.

`dimname` specifies the name of the dimension or questionnaire being analyzed (this name will appear on the graphs).

`noobs` prevents the display of observed points on Item Characteristics Curves (ICCs).

`eap` or `nocorrected` avoids using corrected latent trait estimates (values that best predict the individual scores) on the graphs. Instead, Expected A Posteriori (EAP) estimates are used.

`mle` displays a tabular with the Weighted and Unweighted Most Likely Estimates of the latent trait.

`noresiduals` omits the display of residuals graphs for each item.

`jitter` introduces fluctuations around the position of points on residuals graphs to better represent all individuals. By default, this parameter is set to 0, which means individuals with the same scores are represented by the same point. Increasing this parameter adds more perturbation to the points' locations. The value should be an integer.

`alpha` specifies the confidence level for the intervals shown on residuals graphs.

`wcc` displays the Weighted Category Characteristics Curves (CCCs). In this case, the CCCs are weighted by the density of the latent trait.

`filesave` save the graph files.

`extension` specifies the extension of the graph files (default is png).

`dirsave` specifies the directory where the graph files and the docx file will be saved.

`docx` creates a docx file with the main results and graphs.

`genlt` generates several variables. The new variable *string_eap* contains Expected A Posteriori (EAP) estimates of the latent trait, while *string_se* holds their standard errors. *string_pv* contains simulated Plausible Values, and *string_corr* includes latent trait estimates that best match item-sum scores. *string_ml* and *string_wml* provide (weighted) maximum likelihood estimates obtained by optimizing the probability of each observed item response, with each estimate representing the mean across all observed item answers.

`geninf` provides information associated with each individual.

`replace` indicates that existing files or variables will be replaced.

3.3 Saved results

`pcm` stores the results of the `gsem` in `e()` and the following results in `r()`:

Scalars

r(PSI)	Person Separation Index
r(11)	Marginal Log-Likelihood

Matrices

r(difficulties)	Estimations of the difficulty/threshold parameters.
r(matscorelt)	Estimates of the latent trait associated with each complete response pattern for each item-sum score.
r(matgroupescorelt)	Estimates of the latent trait for each group of individuals used in graphical fit tests.
r(matgroupescoremdlt)	Estimates of the latent trait for each incomplete response pattern within groups used for graphical fit tests.
r(difftest)	Differential Item Functioning (DIF) tests for each item and DIF variable.
r(fit)	OUTFIT and INFIT indices of each item.
r(covariates)	Estimates of the variance of the latent trait and the parameters associated with covariates (predictors of the latent regression).
r(mostlikely)	Estimates of the latent trait for each response category of each item, derived by maximizing the category characteristic curves (CCC). Unweighted values use the CCC directly, while weighted values adjust the CCC by the density function of the latent trait.

4 Identifying the determinants of health literacy: an illustration of how to use the PCM command

We present an analysis of the ECHO study, which involved two independent cross-sectional waves of data collection in France during the Springs of 2020 and 2021 Scarlett et al. (2022). The study focused on interviewing residents of homeless shelters, with our current analysis using data from 2021, encompassing responses from 319 individuals. The aim of this study was to assess the impact of the COVID-19 pandemic in persons experiencing homelessness.

Participants were assessed using the Health Literacy Questionnaire (HLQ) Debussche et al. (2018), which measures an individual's ability to access, understand, and apply health information, interact with healthcare professionals, evaluate information, and make informed decisions to improve their health. For individuals who did not have proficiency in written French, a translator could assist them in completing the questionnaire. This analysis examines two dimensions of the HLQ: "Social Support for Health" and "Navigating the Healthcare System". Here, we focus specifically on the "Social Support for Health" dimension, comprising 5 items. These items evaluate whether individuals believe they have multiple relatives who can understand, support, and accompany them, especially in health-related difficulties. Respondents select an answer among four categories, ranging from "Strongly Disagree" (coded as 1) to "Strongly Agree" (coded as 4). The standard approach to scoring involves calculating the mean response across these five items, where a lower score reflects poorer social support for health.

In this section, we illustrate another method to derive a measure from this questionnaire using a partial credit model. Out of the 319 individuals, 318 answered at least one of the five items. Initially, responses were recoded into a range from 0 to 3. The preliminary analysis indicated that all items were dysfunctional (see, for example,

the Category Characteristic Curve of the first item in Figure 2). Specifically, the second response category ("Disagree" - coded 1) exhibited a low probability of selection. Consequently, we combined categories 0 and 1 ("Strongly Disagree" and "Disagree"). Despite this adjustment, Item 4 continued to be dysfunctioning. Prior to recoding this item, we tested the equality of the gaps between the threshold parameters δ_{j1} and δ_{j2} to assess whether a Rating Scale Model (RSM) would be an appropriate model for this dimension.

```
. recode hlq_soc1-hlq_soc5 (0/1=0) (2=1) (3=2)
. qui pcm hlq_soc1 hlq_soc2 hlq_soc3 hlq_soc4 hlq_soc5 ,rsm
. estimates save RSM,replace
file RSM.ster saved
. qui pcm hlq_soc1 hlq_soc2 hlq_soc3 hlq_soc4 hlq_soc5 ,
. estimate store PCM
. estimates use RSM
. estimate store RSM
. lrtest RSM PCM
Likelihood-ratio test
Assumption: RSM nested within PCM
LR chi2(4) = 18.85
Prob > chi2 = 0.0008
```

Since the global test is significant ($p < 0.001$), indicating that the threshold gaps between items vary, the Rating Scale Model (RSM) is not suitable for this data. To address the issue with Item 4, we combined the response categories "Agree" and "Strongly Agree" for this item only. After this recoding, no items were found to be dysfunctioning. We present the results using the recoded items with the pcm command.

```
. gen hlq_soc4bis=hlq_soc4
(325 missing values generated)
. recode hlq_soc4bis (0=0) (1/2=1)
(116 changes made to hlq_soc4bis)
. label variable hlq_soc1 "Item 1"
. label variable hlq_soc2 "Item 2"
. label variable hlq_soc3 "Item 3"
. label variable hlq_soc4bis "Item 4"
. label variable hlq_soc5 "Item 5"
. pcm hlq_soc1 hlq_soc2 hlq_soc3 hlq_soc4bis hlq_soc5,graph genlt(lt_soc) jitter(5) replace
dimname(HLQ/Social support for health)
Number of excluded individuals: 317 (individuals with no responses to any of the items)
Number of analysed individuals: 318
Number of individuals with a complete response pattern: 280
Number of items: 5
Marginal log-likelihood: -1378.9607
```

Items	<-DIF variables->	Threshold	Estimate	s.e.	z	p	<--95% IC -->	
							Lower Bound	Upper Bound

Item 1	1	0.37	0.17	2.19	0.029	0.04	0.71
	2	0.78	0.20	3.95	0.000	0.39	1.17
Item 2	1	-0.22	0.17	-1.28	0.202	-0.56	0.12
	2	0.75	0.19	4.00	0.000	0.38	1.12
Item 3	1	0.58	0.17	3.39	0.001	0.24	0.92
	2	0.85	0.21	4.12	0.000	0.44	1.25
Item 4	1	-0.37	0.15	-2.41	0.016	-0.67	-0.07
Item 5	1	-0.16	0.17	-0.96	0.338	-0.50	0.17
	2	0.69	0.19	3.74	0.000	0.33	1.06
Variance		1.77	0.27	6.46	0.000	1.23	2.30

Mean squared std error of the latent variable: 0.49
 Global variance of the latent variable: 2.25
 PSI: 0.78

Items	<-DIF variables->		<--- Standardized --->	
	OUTFIT	INFIT	OUTFIT	INFIT
Reference values*	[0.66;1.34]	[0.89;1.11]	[-2.6;2.6]	[-2.6;2.6]
Reference values**	0.75;1.30]	[0.75;1.30]	[-2.6;2.6]	[-2.6;2.6]
Item 1	1.157	1.053	1.314	0.708
Item 2	0.971	1.017	-0.262	0.262
Item 3	0.941	0.948	-0.434	-0.635
Item 4	1.244	1.118	2.315	2.110
Item 5	1.109	1.082	1.105	1.137

*: As suggested by Wright (Smith, 1998)
 **: As suggested by Bond and Fox (2007)

Group	Score	Freq	Latent Trait Mean	Trait s.e.	Expected Score	Corrected Estimate
1 (n=33)	0	33	-1.796	0.860	1.16	-2.997
	0/0	33	-1.797	0.860	1.16	-2.997
2 (n=40)	1	30	-1.155	0.747	1.52	-1.542
	.	10	-1.568	0.903	1.47	-2.479
	1/1	40	-1.259	0.786	1.51	-1.776
3 (n=44)	2	39	-0.657	0.671	2.10	-0.751
	.	5	-0.889	0.788	2.07	-1.119
	2/2	44	-0.683	0.684	2.10	-0.793
4 (n=38)	3	31	-0.241	0.623	2.84	-0.239
	.	7	-0.411	0.702	3.63	-0.449
	3/3	38	-0.272	0.638	2.99	-0.278
5 (n=35)	4	29	0.131	0.598	3.58	0.186
	.	6	0.037	0.657	3.29	0.079
	4/4	35	0.114	0.608	3.53	0.168

6 (n=31)	5	27	0.483	0.592	4.47	0.585	
	.	4	0.487	0.672	5.05	0.589	
	-----		5/5	31	0.483	0.603	4.55
7 (n=34)	6	32	0.840	0.605	5.28	0.997	
	.	2	0.655	0.761	5.96	0.783	
	-----		6/6	34	0.828	0.614	5.32
8 (n=63)	7	26	1.224	0.639	6.54	1.489	
	8	14	1.669	0.700	7.05	2.233	
	9	19	2.226	0.799	7.61	3.629	
	.	4	1.379	0.760	6.77	1.888	
	-----		7/9	63	1.634	0.709	6.99

The first table of results displays the estimated threshold parameters for each item. Notably, Item 4 has only one threshold parameter due to its reduced number of response categories after recoding (now two), whereas the other items have two parameters each.

The Personal Separation Index (PSI) is estimated at 0.78, indicating a satisfactory level of reliability for this dimension.

The second table shows the INFIT and OUTFIT statistics for each item. The OUTFIT values range from 0.941 to 1.244, while the INFIT values range from 0.948 to 1.118. These ranges suggest a good fit of the model to the data.

The third table categorizes individuals into eight groups, each with more than 30 members. For each group, we report the frequencies of scores, the estimated latent trait using Expected A Posteriori (EAP) estimates ("Latent Trait Mean"), and the corrected latent trait estimates ("Corrected Latent Trait").

The final table presents the best estimates for each answer category of each item regarding the latent trait.

The Item Characteristic Curves (ICCs) and Test Characteristic Curves (TCCs) (e.g., Figure 4 and Figure 5) demonstrate a strong graphical fit between the model and the data. The information curves (Figures 6, 7 and 8) reveal that Items 1 and 3 provide information for measuring individuals at higher levels, whereas Item 4 provides information for measuring individuals at lower levels, albeit providing less information due to having only two response categories. Items 2 and 5 offer consistent information across the entire continuum.

The Wright map (Figure 9) illustrates a good alignment between the average level of individuals (maximum of the latent trait density function) and the optimal measurement level of the questionnaire (maximum of the information curve): the density and information curves are maximized at similar levels. The distribution of scores (indicated by the red and light pink bars) approximates a uniform distribution, suggesting a sample with a wide range of latent variable levels (wide range of health literacy associated

to social support).

Finally, we examined the determinants of the "Social Support for Health" latent variable using a latent regression model with the `continuous` and `categorical` options. The variables considered include:

- French speaking level (4 levels: "Not at all," "Not very well," "Pretty good," "Very good")
- French writing level (4 levels: "Not at all," "Not very well," "Pretty good," "Very good")
- French reading level (4 levels: "Not at all," "Not very well," "Pretty good," "Very good")
- Education level (6 levels: "None" to "University level")
- Sex (2 levels: "Male," "Female")
- Partnership status (3 levels: "Lives with partner," "Had a partner but lives alone," "Lives alone")
- Presence of a chronic disease (2 levels: "Yes," "No")
- Health perception (4 levels: "Bad," "Pretty good," "Very good," "Excellent")
- Age (continuous)

All variables were included in the latent regression model. We used a backward elimination procedure to remove non-significant variables at the 5% level. Intermediate tests led to the grouping of some variable levels. The selection procedure is detailed in Table 2.

Main results for the final model are presented here:

```
. label variable parle "French speaking level"
. label variable lecture2 " French reading level"
. label variable sexe2 "Sex"
. label variable sante3 "Health perception"
. pcm hlq_soc1 hlq_soc2 hlq_soc3 hlq_soc4bis hlq_soc5 , cat(parle lecture2 sexe2 sante3 )
Number of excluded individuals: 317 (individuals with no responses to any of the items)
Number of analysed individuals: 318
Number of individuals with a complete response pattern: 280
Number of items: 5
Marginal log-likelihood: -1352.2777
```

Items (output omitted)	<-DIF variables-> Threshold	Estimate	s.e.	z	p	<--95% IC -->	
						Lower Bound	Upper Bound
Variance		1.56	0.25	6.28	0.000	1.07	2.04
French speaking level	1	1.26	0.34	3.72	0.000	0.60	1.93

Modified variable	Modification	p-value	Variance of latent variable	% explained variance
<i>Model without any variable</i>			1.77	0%
<i>Model with all tested variables</i>			1.43	19%
Education Level	dropped	0.52	-	-
French writing level	dropped	0.44	1.49	16%
Partner	dropped	0.23	1.50	15%
Chronic disease	dropped	0.46	1.50	15%
Health perception	"bad" and "pretty good" grouped	0.35	1.51	15%
Health perception	"very good" and "excellent" grouped	0.44	1.52	14%
French speaking level	"pretty good" and "very good" grouped	0.32	1.53	14%
French speaking level	"not very well", "pretty good" and "very good" grouped	0.15	1.54	13%
French reading level	"not at all" and "not very well" grouped	0.79	1.54	13%
Age	continuous → 2 levels " < 40" and " ≥ 40"	0.19	1.55	12%
Age	dropped	0.26	1.55	12%
French reading level	"not at all", "not very well", "very good" grouped	0.76	1.56	12%
<i>Final model</i>			1.56	12%

Table 2: Selection procedure of the variables for the latent regression model

French reading level	1	-0.54	0.21	-2.61	0.009	-0.94	-0.13
Sex	1	-0.43	0.18	-2.37	0.018	-0.78	-0.07
Health perception	1	0.40	0.18	2.22	0.026	0.05	0.75

Mean squared std error of the latent variable: 0.48

Global variance of the latent variable: 2.04

PSI: 0.77 (without adjustment on covariates)

(output omitted)

We interpret these results as follows:

- A latent variable level of "0" corresponds to the average level for a woman who neither speaks nor reads French and rates her health as "bad" or "pretty good".
- Individuals who speak French "not very well," "pretty well," or "very well" have a latent variable level that is 1.26 units [0.60; 1.93] higher than those who do not speak French at all.
- Individuals who read French "pretty well" have a latent variable level that is 0.54 units [0.13; 0.94] lower than those with other reading levels.
- Males have a latent variable level that is 0.43 units [0.07; 0.78] lower than females.
- Individuals who rate their health as "very good" or "excellent" exhibit a higher latent variable level by 0.40 units [0.05; 0.75] compared to others.
- The variables included in the model account for 12% ($= 1 - 1.56/1.77$ using variances of the final and empty models) in the variance in the latent variable.

5 Conclusion

The `pcm` command proves to be both a powerful and sophisticated tool, designed for in-depth data analysis within the framework of Rasch measurement theory. This tool is distinguished by its ability to produce a wide range of tables and graphs, thus providing a detailed and nuanced overview of data from questionnaires or ability tests.

The `pcm` command supports managing data that exhibit differential item functioning (DIF) by adjusting item threshold parameters, fit indices, and latent variable estimates, as illustrated in Touzani et al. (2024). Additionally, the `pcm` command enables linking and equating between two sets of items, as demonstrated in Hardouin et al. (2024). This functionality facilitates the comparison and integration of different item sets, ensuring consistency and equivalence across diverse measurement instruments. Nevertheless, for the sake of parsimony, we have chosen not to present these two functionalities in this article. Instead, we have focused on parameter estimation, model fit, various graphical representations and latent regression to avoid overloading this paper.

The core features of the `pcm` command are seamlessly integrated into the Pro-Online software platform (<https://pro-online.net/>), which operates under Numerics by Stata licenses.

6 References

- Andrich, D. 1978. A rating formulation for ordered response categories. *Psychometrika* 43: 561–573.
- Andrich, D., and I. Marais. 2019. *A Course in Rasch Measurement Theory : Measuring in the Educational, Social and Health Sciences*. Springer.
- Bond, T., and C. M. Fox. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Routledge.
- Christensen, K. B., S. Kreisner, and M. Mesbah. 2012. *Rasch models in health*. John Wiley and Sons, Ltd.
- Debussche, X., V. Lenclume, M. Balcou-Debussche, D. Alakian, C. Sokolowsky, D. Ballet, G. R. Elsworth, R. H. Osborne, and L. Huiart. 2018. Characterisation of health literacy strengths and weaknesses among people at metabolic and cardiovascular risk: Validity testing of the Health Literacy Questionnaire. *SAGE Open Med* 21(6): 2050312118801250.
- Fischer, G. H., and I. W. Molenaar. 1995. *Rasch models: Foundations, recent developments and applications*. Springer.
- Hamel, J.-F., V. Sébille, G. Challet-Bouju, and J.-B. Hardouin. 2016. Partial Credit Model: Estimations and Tests of Fit with Pcmmodel. *The Stata Journal* 16(2): 464–481.
- Hamel, J.-F., V. Sébille, T. le Néel, G. Kubis, F.-C. Boyer, and J.-B. Hardouin. 2017. What are the appropriate methods for analyzing patient-reported outcomes in randomized trials when data are missing? *Statistical Method in Medical Research* 26(6): 2897–2908.
- Hardouin, J.-B. 2007. Rasch Analysis: Estimation and Tests with Raschtest. *The Stata Journal* 7(1): 22–44.
- Hardouin, J.-B., R. Conroy, and V. Sébille. 2011. Imputation by the mean score should be avoided when validating a Patient Reported Outcomes questionnaire by a Rasch model in presence of informative missing data. *BMC Medical Research Methodology* 11: 105.
- Hardouin, J.-B., J. Coste, A. Leplège, and A. Rouquette. 2024. Equating and linking Patient-Reported Outcomes Measurement Information System 29-item questionnaire and 36-item Short-Form Health Survey domains using Rasch modeling. *J Clin Epidemiol* 169: 111326.
- Kondratek, B. 2022. uirt: A command for unidimensional IRT modeling. *The Stata Journal* 22(2): 243–268.
- Linacre, J. M. 2022. What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions* 16(2).

- Masters, G. 1982. A Rasch model for partial credit scoring. *Psychometrika* 47: 149–174.
- Müller, M. 2020. Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications* 7(5).
- Rasch, G. 1980. *Probabilistic models for some intelligence and attainment tests*. The University of Chicago Press.
- Scarlett, H., M. Melchior, C. Davaisse-Paturet, T. E. Aarbaoui, C. Longchamps, N. Figueiredo, and S. Ducarroz. 2022. Substance use among residents of homeless shelters during the COVID-19 Pandemic: findings from France. *International Journal of Public Health* 67: 1604684.
- Smith, R., R. Schumacker, and M. Bush. 1998. Using item mean squares to evaluate fit to the Rasch model. *J Outcome Meas* 2(1): 66–78.
- Touzani, R., A. Rouquette, E. Schultz, C. Allaire, P. Carrieri, J. Mancini, and J.-B. Hardouin. 2024. Psychometric validation of the French version of two scales measuring general (HLS19-Q12) and navigational (HLS19-NAV) health literacy using the Rasch model. *BMC Public Health* 24: 3079.
- Wright, B., and M. Stone. 1999. *Measurement essentials*. Wide Range, Inc.
- Wright, B. D. 1998. Estimating Rasch measures for extreme scores. *Rasch Measurement Transaction* 12(2): 632–633.
- Zheng, X., and S. Rabe-Hesketh. 2007. Estimating Parameters of Dichotomous and Ordinal Item Response Models with Gllamm. *The Stata Journal* 7(3): 313–333.

About the authors

Jean-Benoit Hardouin is an associate professor of Clinical Research, Biostatistics, and Epidemiology at Nantes University and the University Hospital of Nantes. His research focuses on applying latent variable psychometric models in clinical research and epidemiology, with a particular emphasis on measuring health literacy.

Yseulys Dubuy is an associate professor of Biostatistics at Nantes University. Her research centers on psychometrics, specifically in detecting response shift.

Myriam Blanchin is a researcher in Biostatistics at Nantes University. Her research also focuses on psychometrics, notably in response shift detection.

Simon Ducarroz is a researcher in Public Health, with a particular focus on social epidemiology.