

Supplemental Material for

NEPS*scaling*: Providing Plausible Values for Users of the National Educational Panel Study

Further Functionality of NEPS*scaling* Shiny App 2

Further Functionality of NEPS*scaling* Shiny App

To import background data for the estimation, a click on the field “Manage background data” on the top left of the starting page (also called the “Manage” tab) expands the import, remove and inspect options for the background data (Figure S1). By clicking on the “Browse...” button, the file explorer pops up and the background data can be selected. After setting the scale level of the background variables, the data is displayed in the “Manage” tab (Figure S2).

Furthermore, the background data can be inspected, i.e., there are options to select variables and filter columns available after expanding the “Manage background data” field’s “Inspect background data” button (see Figure S3). Please note that these configurations do not affect the variable selection for the plausible values estimation.

Figure S1

Import of background data

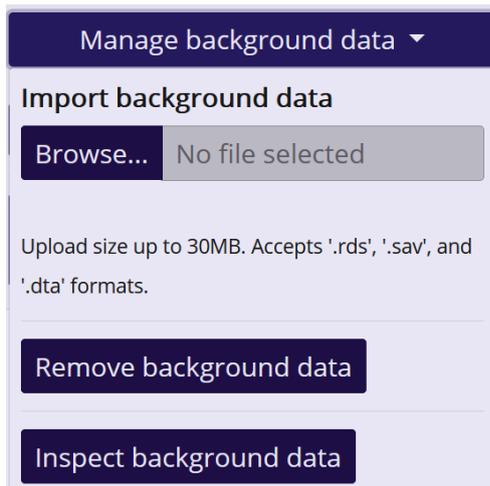
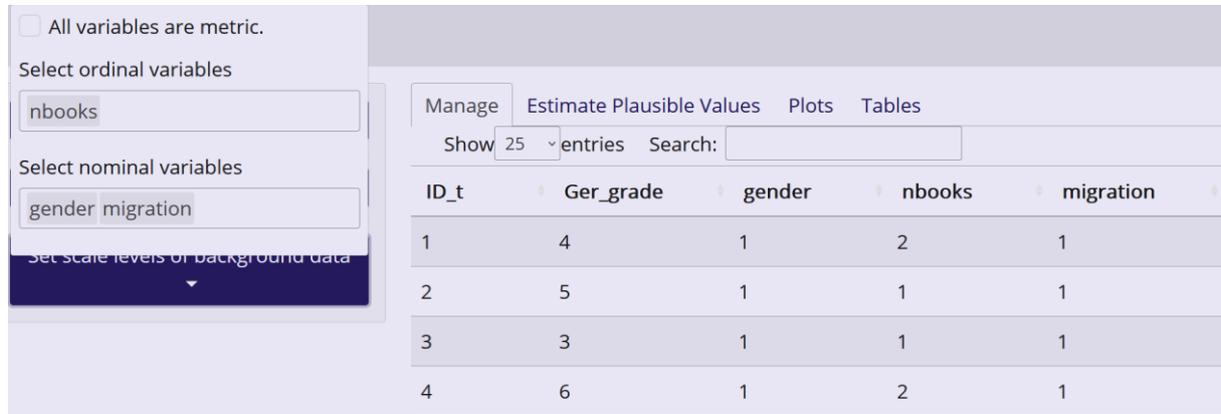


Figure S2

Display of background data after setting the scale level



The screenshot shows a software interface with a sidebar on the left and a main data table on the right. The sidebar contains the following elements:

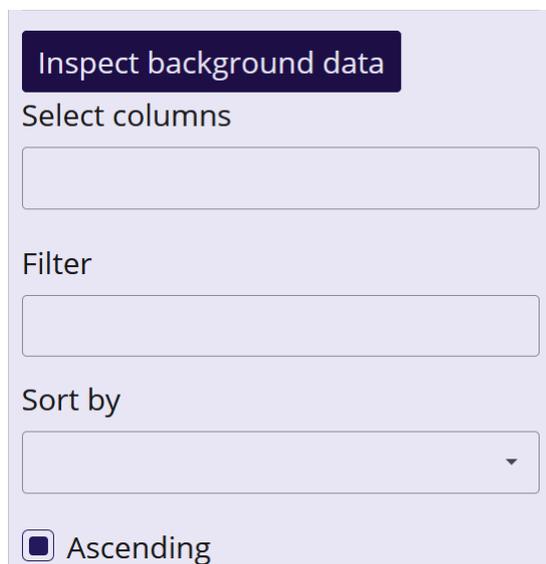
- All variables are metric.
- Select ordinal variables:
- Select nominal variables:
- Set scale levels of background data (highlighted in dark blue)

The main table has a header with tabs: Manage, Estimate Plausible Values, Plots, Tables. Below the tabs, it shows "Show 25 entries" and a search box. The table data is as follows:

ID_t	Ger_grade	gender	nbooks	migration
1	4	1	2	1
2	5	1	1	1
3	3	1	1	1
4	6	1	2	1

Figure S3

Ways of inspecting the background data



The screenshot shows a software interface for inspecting background data. It includes the following elements:

- Inspect background data** (highlighted in dark blue)
- Select columns:
- Filter:
- Sort by:
- Ascending

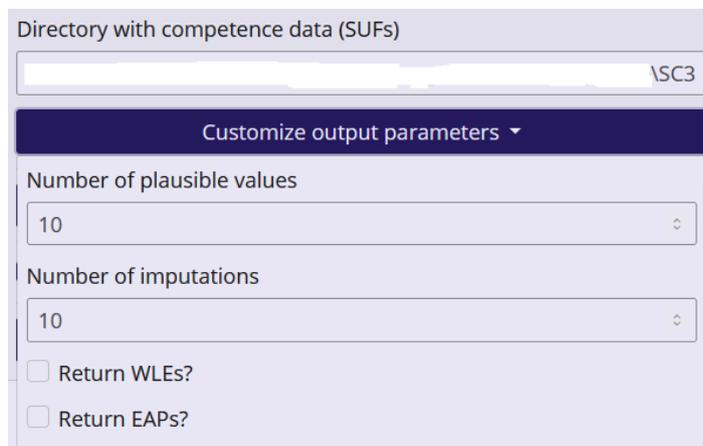
Note. Any variable selection or subject filtering is not permanent. Exclusion of variables for the plausible values estimation is part of the input arguments for the estimation.

Before estimating plausible values, the output of the estimation algorithm can be tweaked. After expanding the field “Customize output parameters” in the middle left of the “Estimate Plausible Values” tab (Figure S4), one can change the number of imputations and plausible values (both default to 10) and whether point estimates of the competences (weighted likelihood estimates, WLEs, and expected a posteriori estimates, EAPs) should be returned next to the set of plausible values.

The field “Customize model parameters” includes the option to exclude background variables from the plausible values estimation (i.e., they are used only for imputation). The variables can be chosen from a drop down selection of the available variables (Figure S5).

Figure S4

Customization of the output object



Directory with competence data (SUFs)

\SC3

Customize output parameters ▾

Number of plausible values

10

Number of imputations

10

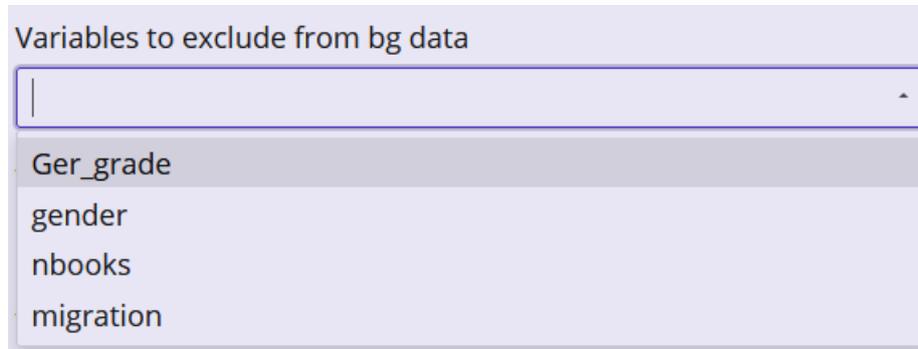
Return WLEs?

Return EAPs?

Note. The pre-defined numbers of imputations and plausible values are the current standard in large-scale assessment studies and should be considered as a minimum number.

Figure S5

Exclusion of variables from the background data during plausible values estimation



Variables to exclude from bg data

- Ger_grade
- gender
- nbooks
- migration

Note. This field is part of the “customize model parameters” button. It refers to either the background data in cross-sectional estimation or the background data of the first wave in longitudinal estimation. For later waves, separate fields appear after selecting longitudinal estimation.

After the estimation, the imputed data and plausible values can be inspected visually in the “Plots” tab (Figure S6). There are options to display the imputed data and plausible values with histograms, scatter plots and kernel density plots (Figures S7 and S8). The variable to plot on the x-axis and, in case of the scatter plot, the variable to plot against on the y-axis can be chosen from drop down menus. Furthermore, the data can be color coded by group memberships of the participants (see Figure S8).

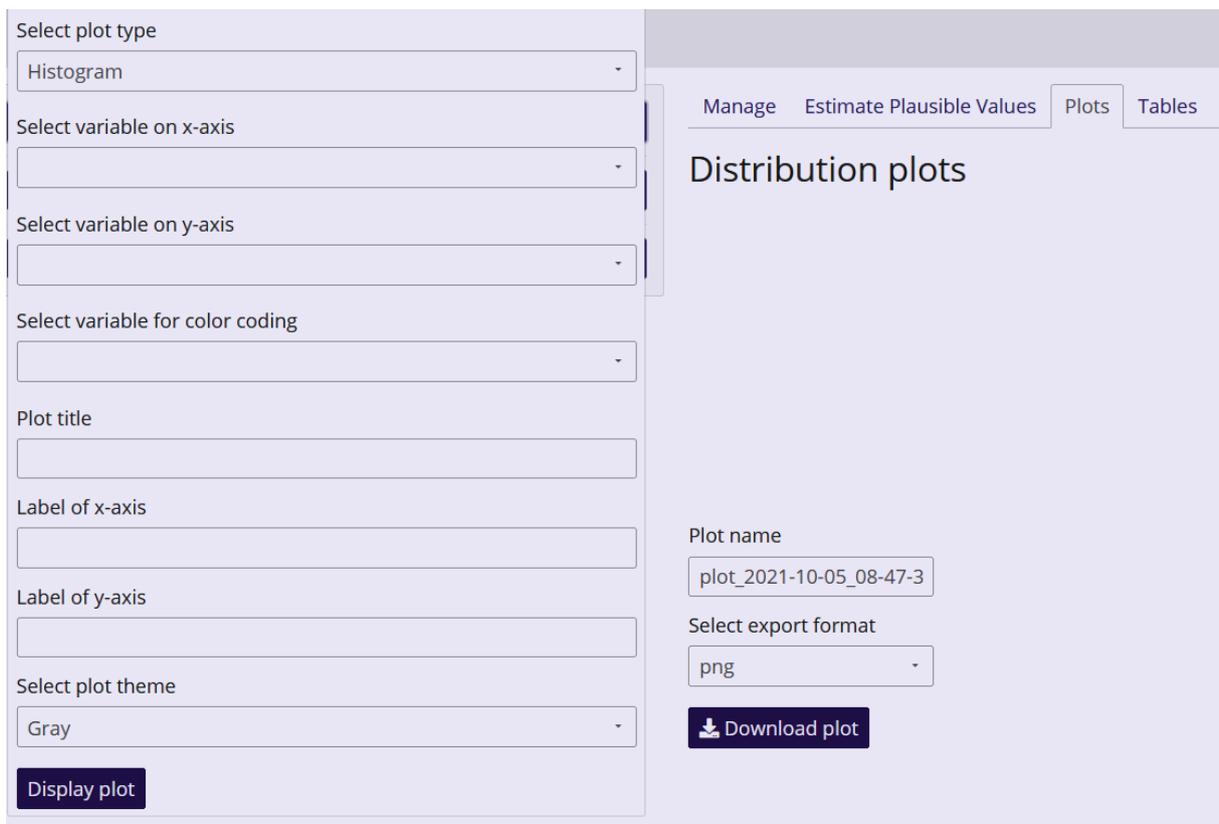
Figure S6

Overview of the Plots tab



Figure S7

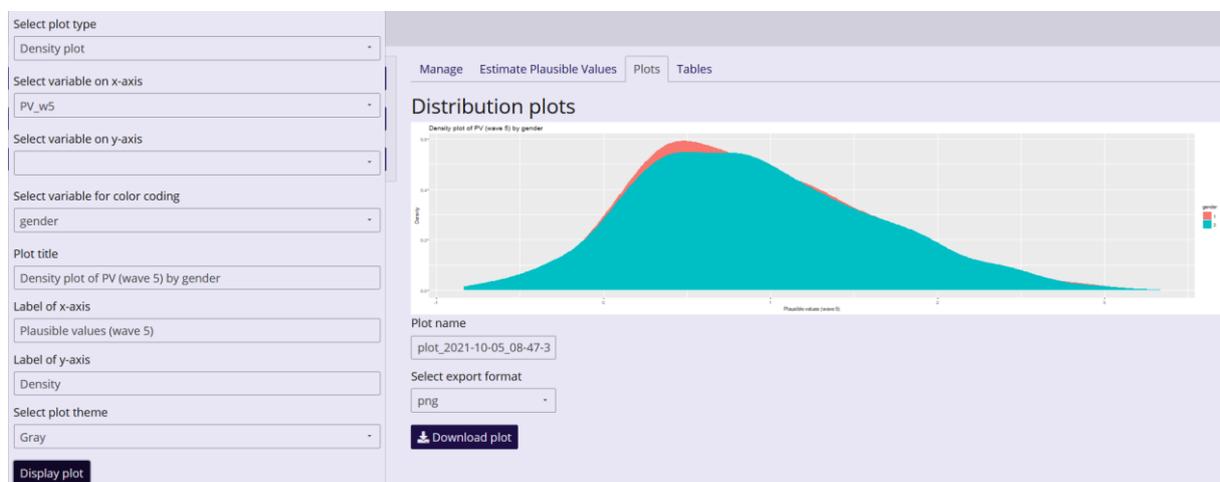
Options for plotting the imputed data and estimated plausible values



Note. Histograms, scatter plots and kernel density plots are available for the imputed background data and the plausible values, aggregated to one single data set. *Color coding* refers to group-wise plots, e.g., grouped by gender.

Figure S8

Example for kernel density plot of plausible value by gender



Moreover, the imputation process can be inspected as well (fields “Imputation tree structures” and “Variable importance plots for imputations” in the “Plots” tab; Figure S6). Both plots need the same inputs as can be seen in Figures S9 and S10, that is, the imputation and the imputed variable which can be selected from drop down menus in both cases. However, the tree structure may become very complex especially if many variables are included in the model. In those cases, the tree structure may not be displayed by NEPSshiny. It can still be inspected in its written representation using the function `get_imputation_tree(pv_obj, imputation, variable)` in R. The variable importance plots are not affected by the tree’s complexity.

Figure S9

Imputation tree plot

The screenshot shows a web interface with a sidebar on the left and a main content area on the right. The sidebar has two main sections: 'Plots for plausible values and imputations' and 'Imputation tree structures'. Under 'Imputation tree structures', there are two dropdown menus: 'Select imputation' and 'Select variable', followed by a 'Display tree plot' button. The main content area has a navigation bar with 'Manage', 'Estimate Plausible Values', 'Plots', and 'Tables'. Below the navigation bar, the title 'Imputation tree plots' is displayed. There are two input fields: 'Plot name' with the value 'cart_2021-10-05_08-47-3' and 'Select export format' with the value 'png'. A 'Download plot' button is located at the bottom right of the main content area.

Note. For each variable per imputation, the classification or regression tree used for filling in the missing values can be displayed, as long as the tree is not too complex.

Figure S10

Variable importance plot

The screenshot shows a web application interface. On the left, a sidebar contains three dark blue buttons with white text: "Plots for plausible values and imputations", "Imputation tree structures", and "Variable importance plots for imputations". The main content area has a top navigation bar with tabs: "Manage", "Estimate Plausible Values", "Plots" (which is highlighted), and "Tables". Below the tabs, the heading "Variable importance plots" is displayed. Underneath, there are two dropdown menus labeled "Select imputation" and "Select variable". Below these is a dark blue button labeled "Display variable importance plot". To the right of the main content area, there is a "Plot name" input field containing the text "variable_importance_20:", a "Select export format" dropdown menu with "png" selected, and a dark blue button labeled "Download plot" with a download icon.

Note. For each variable per imputation, the importance of the predictor variables in the imputation can be plotted as bar plots.

The imputed data sets can, furthermore, be inspected in tabular form. By clicking the button “Descriptive tables for plausible values and imputations” in the “Tables” tabs, descriptive statistics (e.g., mean, standard deviation and range) of the aggregated imputed data sets, including the estimated plausible values, are calculated and displayed (Figure S11).

Figure S11

Descriptive table of imputed data sets

Descriptive table

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ID_t	3000.00	1500.50	866.17	1500.50	1500.50	1111.95	1.00	3000.00	2999.00	0.00	-1.20	15.81
Ger_grade	3000.00	3.48	1.42	3.00	3.48	1.48	1.00	6.00	5.00	0.00	-0.76	0.03
gender*	3000.00	1.51	0.50	2.00	1.51	0.00	1.00	2.00	1.00	-0.03	-2.00	0.01
nbooks*	0.00	NA	NA	NA	NA	NA	Inf	-Inf	NA	NA	NA	NA
migration*	3000.00	1.15	0.35	1.00	1.06	0.00	1.00	2.00	1.00	1.99	1.97	0.01
mag5_sc1u_schavg	3000.00	-0.01	0.13	0.01	-0.02	0.15	-0.23	0.36	0.60	0.49	0.10	0.00
mag7_sc1u_schavg	3000.00	-0.01	0.13	0.01	-0.02	0.15	-0.23	0.36	0.60	0.49	0.10	0.00
mag9_sc1u_schavg	3000.00	-0.01	0.13	0.01	-0.02	0.15	-0.23	0.36	0.60	0.49	0.10	0.00
mag12_sc1u_schavg	3000.00	-0.01	0.13	0.01	-0.02	0.15	-0.23	0.36	0.60	0.49	0.10	0.00
PV_w1	3000.00	-0.61	0.79	-0.64	-0.62	0.82	-3.08	1.80	4.88	0.12	-0.36	0.01
PV_w3	3000.00	0.12	0.69	0.12	0.11	0.72	-1.84	2.37	4.21	0.07	-0.16	0.01
PV_w5	3000.00	0.91	0.70	0.83	0.88	0.71	-0.83	3.33	4.17	0.44	-0.16	0.01
PV_w9	3000.00	1.51	0.56	1.54	1.52	0.50	-0.23	3.30	3.54	-0.28	0.19	0.01

Descriptive Statistics of Average Imputed Data Sets. * Factor variables.

Table name
descriptives_2021-10-05

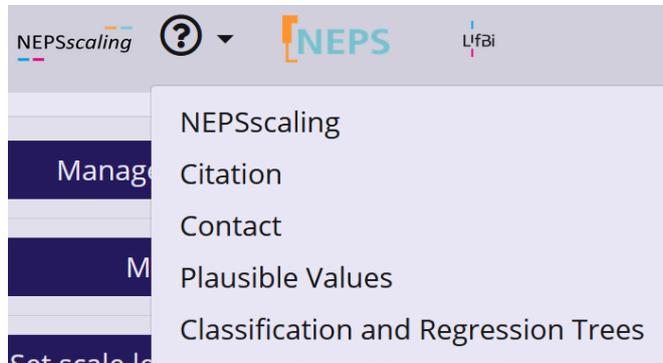
Download Descriptives

Note. The descriptive table is created for the aggregated completed data sets, including the estimated plausible values.

Finally, help and support information is available when clicking on the question mark at the top left of the Shiny app (see Figure S12). It links to a short summary of the purpose of *NEPSscaling* (option “NEPSscaling”; see also Figure S13), citation (option “Citation”; see Figure S14) and contact information in case of further questions (option “Contact”; see also Figure S15) as well as an overview of the theory behind plausible values (option “Plausible Values”; see also Figure S16) and classification and regression trees (“Classification and Regression Trees”; see also Figure S17) that are used for multiple imputation in *NEPSscaling*.

Figure S12

Help and support button in the main bar of the app



Note. A click on the question mark button opens a list of short explanations about the underlying concepts as well as further information (sources).

Figure S13

Short summary of NEPSscaling

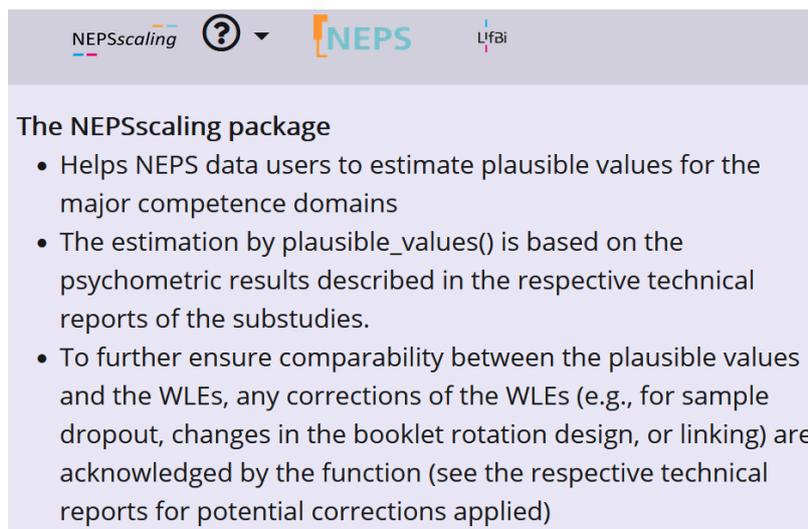
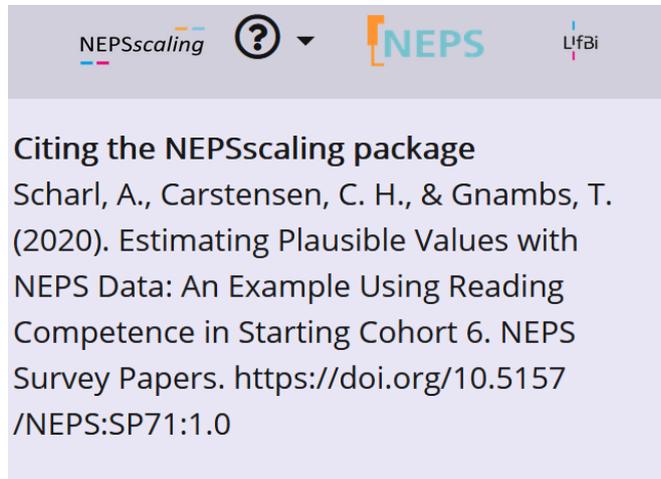


Figure S14

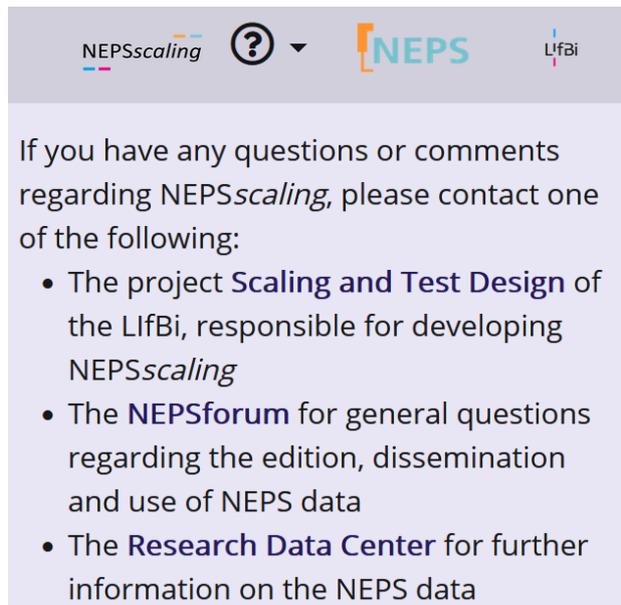
Correct citation of the NEPSscaling package



The screenshot shows the top navigation bar of the NEPSscaling website with logos for NEPSscaling, a help icon, NEPS, and LfBi. Below the navigation bar, the text reads: "Citing the NEPSscaling package" followed by the citation: "Scharl, A., Carstensen, C. H., & Gnamb, T. (2020). Estimating Plausible Values with NEPS Data: An Example Using Reading Competence in Starting Cohort 6. NEPS Survey Papers. <https://doi.org/10.5157/NEPS:SP71:1.0>"

Figure S15

Contact information of NEPSscaling team



The screenshot shows the top navigation bar of the NEPSscaling website with logos for NEPSscaling, a help icon, NEPS, and LfBi. Below the navigation bar, the text reads: "If you have any questions or comments regarding NEPSscaling, please contact one of the following:" followed by a bulleted list of contact options:

- The project **Scaling and Test Design** of the LfBi, responsible for developing NEPSscaling
- The **NEPSforum** for general questions regarding the edition, dissemination and use of NEPS data
- The **Research Data Center** for further information on the NEPS data

Note. The phrases in bold type contain hyperlinks of the described contacts.

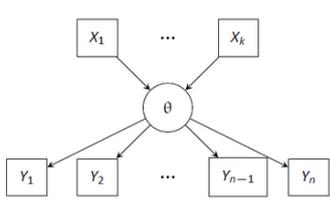
Figure S16

Short summary of plausible values technique with recommended reading

NEPSscaling ? NEPS IZBI

Plausible Values

- Estimators for latent constructs such as competencies
- Set of random draws out of individual respondent's latent competence distribution
- Derived from competence test and respondent characteristics (e.g., gender, socio-economic status)
- Uncertainty in random draws reflects uncertainty in competence estimation
- Background variables should at least contain all variables used for later analysis
- Unbiased on a population level, but biased on the individual level because of respondent information (i.e., group-level information)
- Special case of multiple imputation: statistical analyses with plausible values have to be performed accordingly



The diagram is a path model. At the top, there are boxes for observed variables X_1 , an ellipsis, and X_k . Arrows point from each of these boxes to a central circle containing the Greek letter θ . From the central circle, arrows point down to a row of boxes containing observed variables Y_1 , Y_2 , an ellipsis, Y_{n-1} , and Y_n .

Recommended Reading

1. Scharl, A., Carstensen, C.H., & Gnams, T. (2020). Estimating Plausible Values with NEPS Data: An Example Using Reading Competence in Starting Cohort 6 (NEPS Survey Paper No. 71). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. doi:10.5157/NEPS:SP71:1.0
2. von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are

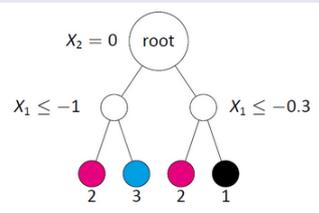
Figure S17

Short summary of used imputation algorithm with recommended reading

NEPScaling ?NEPSLFBI

Classification and Regression Trees

- Background variables for plausible values cannot contain missingness, but non-response is pervasive in large scale assessments and surveys
- Multiple imputation as an approach to fill in randomly missing data without introducing further bias
- Decision trees (e.g., classification and regression trees, CART) can be used to identify a set of plausible responses for the missing data
- Variable with missingness is recursively split into subsets; each subset has to be more homogenous than the superset
- Splits are made according to a value on a predictor variable (e.g., being female, being older than X years) until a node purity criterion is reached
- Prediction for missing values are drawn by following the tree's branches to its nodes and choosing a value from the node following an algorithm
- CART is a non-parametric approach and automatically incorporates non-linear relationships in the predicted and predictor variables



```
graph TD; root((root)) -- "X2 = 0" --> L(( )); root -- "X2 = 0" --> R(( )); L -- "X1 <= -1" --> L2((2)); L -- "X1 <= -1" --> L3((3)); R -- "X1 <= -0.3" --> R2((2)); R -- "X1 <= -0.3" --> R1((1));
```

Recommended Reading

1. Scharl, A., Carstensen, C.H., & Gnamb, T. (2020). Estimating Plausible Values with NEPS