

Additional Study:  
Baden-Wuerttemberg (BW)  
SUF Version 3.0.0  
Data Manual

*Markus Zielonka, Tobias Koberg, Robert Lipp, and Martin Gunsenheimer-Welsch*



Copyrighted Material  
Leibniz Institute for Educational Trajectories (LifBi)  
Wilhelmsplatz 3, 96047 Bamberg  
Director: Prof. Dr. Hans-Günther Roßbach  
Executive Director of Research: Dr. Jutta von Maurice  
Executive Director of Administration: Dr. Robert Polgar  
Bamberg, 2015

## Research Data Papers

at the NEPS Research Data Center, Bamberg

The NEPS Research Data Paper series presents documentation resources prepared to support the work with data from the National Educational Panel Study (NEPS).

Full citation of this document:

Zielonka, M., Koberg, T., Lipp, R., & Gunsenheimer-Welsch, M. (2015, March). *Additional Study: Baden-Wuerttemberg (BW) SUF Version 3.0.0 Data Manual*. NEPS Research Data Paper. University of Bamberg

*This release of scientific use data from the NEPS Additional Study I – “G8 Reform Study in Baden-Wuerttemberg” was prepared by the staff of the NEPS Data Center in tight collaboration with colleagues from the NEPS-Methods Group (weighting) and the staff of NEPS Pillar 1 and NEPS Stage 5 (for scoring and scaling of competencies). It represents a major collective effort. The contribution of the following staff members of the NEPS is gratefully acknowledged:*

### Data preparing, editing and scoring

Daniel Bela (integration of metadata, data preparation tools in Stata)

Christoph Duchhardt (scoring and scaling of Math competencies)

Martin Gunsenheimer-Welsch (editing and integration of 3<sup>rd</sup> wave)

Tobias Koberg (anonymisation, edition of data)

Robert Lipp (editing and integration of 3<sup>rd</sup> wave)

Manuel Munz (coding and classification)

Christin Schanz (examples)

Benno Schönberger (weighting)

Mihaela Tudose (management and managing of metadata)

Wolfgang Wagner (scoring and scaling of all other competencies)

Markus Zielonka (edition and reintegration of data)

### Data manual

Markus Zielonka

Tobias Koberg

Robert Lipp

Martin Gunsenheimer-Welsch

National Educational Panel Study (NEPS)

Research Data Center

Wilhelmsplatz 3

96047 Bamberg, Germany

E-mail: [fdz@lifbi.de](mailto:fdz@lifbi.de)

Web: <https://www.neps-data.de/en-us/datacenter>

Phone: +49 951 8633511



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	About this manual . . . . .	1
1.2	Obtaining the data . . . . .	1
1.3	Three modes of data access . . . . .	2
1.4	Publications with NEPS data . . . . .	2
<b>2</b>	<b>General conventions</b>	<b>3</b>
2.1	File names . . . . .	3
2.2	Variable names . . . . .	3
2.2.1	Special conventions for variables in test data . . . . .	6
2.3	Missing values . . . . .	6
<b>3</b>	<b>Sampling and surveying procedures</b>	<b>8</b>
3.1	Overview . . . . .	8
3.2	Sampling and response rates . . . . .	9
3.3	Competence testing and students questionnaire . . . . .	9
3.4	School heads and subject heads questionnaire . . . . .	10
3.5	School grades and tracking information . . . . .	10
<b>4</b>	<b>Datfiles</b>	<b>11</b>
4.1	Pooled cross sectional target file: xTarget . . . . .	11
4.2	Pooled cross sectional competencies file: xTargetCompetencies . . . . .	11
4.3	Pooled cross sectional file for teachers and heads: xInstitution . . . . .	12
4.4	Linking and method file: Profile . . . . .	12
4.5	Clustering and merging within a multilevel data structure . . . . .	13
<b>5</b>	<b>Generated variables and weights</b>	<b>14</b>
5.1	Coding . . . . .	14
5.2	Weights . . . . .	14
<b>6</b>	<b>Anonymization</b>	<b>15</b>
6.1	Specifications . . . . .	15
6.2	Onion-shaped model . . . . .	15
6.3	Conducted measures . . . . .	17
<b>7</b>	<b>Examples</b>	<b>18</b>
7.1	Example 1—Merging data from xInstitution and xTarget via Profile . . . . .	18
7.2	Example 2—Using Profile . . . . .	20
<b>8</b>	<b>Tools for Stata users</b>	<b>22</b>
8.1	Multi-lingual data sets . . . . .	22
8.2	Data signatures . . . . .	22
8.3	NEPStools . . . . .	22
8.3.1	Installation and updating . . . . .	23
8.3.2	<i>nepsmiss</i> —Recoding missing values . . . . .	23
8.3.3	<i>infoquery</i> —Display additional metadata . . . . .	24
8.3.4	<i>charren</i> —Rename variables to survey names . . . . .	24
<b>9</b>	<b>Further information</b>	<b>25</b>

<b>A Appendix</b>	<b>26</b>
A.1 Occupational coding . . . . .	26
A.2 Variables affected by anonymization methods . . . . .	29
<b>References</b>	<b>30</b>

## List of Tables

1	Naming conventions of file names . . . . .	5
2	Overview of missing codes . . . . .	7
3	Overview of samples and survey instrument participation . . . . .	9
4	Different general perspectives and possible levels/clusters within the data . . . . .	13
5	Overview for coded variables (German labels also provided) . . . . .	14
6	Availability of sensitive data . . . . .	17
7	Data set comparison before and after using <i>nepsmiss</i> . . . . .	23
8	Coding of BLK from (Blossfeld, 1985) . . . . .	26
9	Coding of ISCED-97 (United Nations Educational, Scientific and Cultural Organization (UNESCO), 1997) . . . . .	26
10	Coding of CASMIN . . . . .	27

## List of Figures

1	General procedure on students' competence/ability tests and surveying in school	10
2	Recommended merging relation for the datafiles . . . . .	13
3	Onion-shaped model defining the different anonymization levels . . . . .	16

## List of syntax examples

1	Merging information from xInstitution and xTarget data via Profile . . .	18
1	Merging information from xInstitution and xTarget data via Profile . . .	19
2	Using Profile . . . . .	20
2	Using Profile . . . . .	21

# 1 Introduction

## 1.1 About this manual

This manual is intended to assist your work with the data of the NEPS additional study *G8 Reform Study in Baden-Wuerttemberg* (NEPS BW 3.0.0). We aim at providing a guide of how to use these data for your research. Therefore, our focus is on practical aspects of data usage such as the dataset structure, key variables, and examples of data retrievals.

This manual is not a comprehensive documentation resource. Please consult our website <sup>1</sup> for background information on the studies, survey instruments, a structured documentation, and many more resources.

Our goal is to keep this manual as short and simple as possible. At several places, we reference supplementary documents presenting additional information that we consider essential for working with our data:

- Codebook
- Technical reports/working papers on:
  - Weighting (Schönberger, Aßmann, & Zinn, 2015)
  - Scaling of Math competencies (forthcoming)
  - Scaling of Physics, Biology and English competencies (forthcoming)

You can download these documents here:

→ [www.neps-data.de](http://www.neps-data.de) > Data Center > Research Data > Additional Study Baden-Wuerttemberg

We welcome feedback from our users that will help us improve the quality of this manual and our data for future releases. Please report any feedback to:

- [fdz@lifbi.de](mailto:fdz@lifbi.de)

## 1.2 Obtaining the data

There are three simple steps to obtain the data of this release:

- Sign the data use agreement send it per postal mail to us:
  - [www.neps-data.de](http://www.neps-data.de) > Data Center > Data Access > Data Use Agreements
- After approval, sign in as a registered NEPS user at the login at <https://www.neps-data.de>
- Access the data via one of our three access modes (see below)

Depending on which access mode(s) you choose, you will find all further instructions required to obtain the data on our website.

<sup>1</sup> <https://www.neps-data.de/en-us/datacenter>

## 1.3 Three modes of data access

### 1.3 Three modes of data access

Once you have signed a data use contract with the NEPS, we provide you with three modes of access:

- Download from our website (after login),
- RemoteNEPS (browser-based remote access),
- and on-site access.

These three solutions are designed to support the full range of users' interests and maximize data utility while complying with strict standards of confidentiality protection. Access via *RemoteNEPS* works with *biometrical authentication* and requires *at least one participation in the user training courses* provided by the LIfBi Research Data Center.

#### Sensitive data

Each access mode corresponds to a specific level of data sensitivity. Files that are offered for download include data with the highest level of anonymization. These data are available to registered users from the web portal via a secure connection. Files offered via RemoteNEPS contain more sensitive data within a controlled environment. The analysis of information in high resolution (e. g., fine-grained regional information) is only provided on-site in Bamberg where these data are available within a secure site. For details on the access modes and anonymization techniques, see section 6 on page 15; also, find additional information on our website at

→ [www.neps-data.de](http://www.neps-data.de) > Data Center > Data Access

#### File Format

All files are available in Stata and SPSS format with bilingual variable labels and value labels (German and English). Data stored in Stata format contain both languages within one file (see section 8 for a tutorial on how to switch label language in Stata). SPSS files are provided by each language separately.

### 1.4 Publications with NEPS data

If you publish with NEPS data, it is mandatory to quote the following reference:

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special issue]. *Zeitschrift für Erziehungswissenschaft*, 14. Wiesbaden: VS Verlag für Sozialwissenschaften.

In addition, publications using data from this release must include the following acknowledgment:

*This paper uses data from the National Educational Panel Study (NEPS): G8 Reform Study in Baden-Wuerttemberg, doi 10.5157/NEPS:BW:3.0.0. The NEPS data collection is part of the Framework Programme for the Promotion of Empirical Educational Research, funded by the German Federal Ministry of Education and Research and supported by the Federal States.*

A digital object identifier (DOI) uniquely identifies each release of NEPS data (cf. Wenzig, 2012). The DOI of this release redirects to a landing page providing basic information on the data:

# 2 General conventions

## 2.1 File names

The naming of data files included in this release follow a number of conventions which are summarized in table 1 on page 5.

To give an example, the physical file `BW_xTarget_D_3-0-0.dta` refers to the *Download-version (D)* for the data file *xTarget* of *G8 Reform Study in Baden-Wuerttemberg (BW)* of data release *3.0.0*.

[htp]

## 2.2 Variable names

The G8 Reform Study in Baden-Wuerttemberg collects data of three cross-sectional surveys and tests in one federal state and focuses on a very specific institutional change in Germany. Hence, contrary to the common NEPS naming convention of variables we sometimes provide variable names derived from German abbreviations of the questions or the items in focus. We adopted the common NEPS naming convention only for the first digit, generated variables and variables from competence tests<sup>2</sup>

The first digit indicates to which *primary respondent type* the variable refers; in case of the G8 Reform Study in Baden-Wuerttemberg this character can be “t” (target person), “g”, “e” and “m” (subject heads in German, English or Maths in the institution of the target person) and finally “h” (heads of the institution of the target person).

Additionally, some information about the study structure and administration details are provided by the test coordinator at the schools. This information is sometimes essential to understand the data structure (e. g., which rotation or version of a test was administered for student x etc.) and therefore also included into the specific data sets of the primary respondents. These variables are treated as generated variables and get an “x” as a second digit after the “t”, “e” or “h”.

Optional Suffixes are separated from the previous characters by an underscore. There are three types of suffixes:

### ■ Suffixes for generated variables

Generated variables are indicated by the suffix `_g#` (`_g1`, `_g2`, etc.). In most cases, the running number after `_g` is a simple enumerator. However, there are two types of generated variables that assign meanings to these running numbers: occupational/prestige codes and scales.

<sup>2</sup> Since all competence tests in the school and kindergarten cohorts are conducted with the students/children only, the prefix “t” (for *target*) is omitted and the variables start with a domain-specific abbreviation (see below).



## 2.2 Variable names

### ■ Occupational/prestige codes

g1	=	KldB 1988 (German Classification of Occupations 1988 from (Bundesamt, 1992))
g2	=	KldB 2010 (German Classification of Occupations 2010 from (Bundesagentur für Arbeit, 2011))
g3	=	ISCO-88 (Internat. Standard Classification of Occupations 1988 from (Office, 1990))
g4	=	ISCO-08 (Internat. Standard Classification of Occupations 2008 from (Office, 2012))
g5	=	ISEI-88 (Internat. Socio-Economic Index of Occupational Status 1988 from (H. B. G. Ganzeboom, de Graaf, Treiman, & de Leeuw, 1992))
g6	=	SIOPS-88 (Standard International Occupational Prestige Scale 1988 from (Treiman, 1977))
g7	=	MPS (Magnitude Prestige Scale from (Wegener, 1985))
g9	=	BLK (Blossfeld's Occupational Classification from (Blossfeld, 1985))
g14	=	ISEI-08 (Internat. Socio-Economic Index of Occupational Status 2008 from (H. B. G. Ganzeboom, 2010))
g15	=	CAMSIS (Social Interaction and Stratification Scale from (Müller, Lüttinger, König, & Karle, 1989))
g16	=	SIOPS-08 (Standard International Occupational Prestige Scale 2008 from (H. B. Ganzeboom & Treiman, 2010))

### ■ Scale suffix

As scales are generated by a set of other variables, they are also indicated by the above mentioned nomenclature. For the sake of completeness and clarity, it has to be stated that scales are named according to the first variable of the sequence they were generated from. Their running numbers are in so far meaningful as they count up if and only if the first variable of two scales had been identical.

### ■ Wide-format suffix

Wide-format variables stored are indicated by the suffix *\_w#* (e. g., *\_w1*, *\_w2*, etc.). Note that the wide-suffix not necessarily implies a wave logic. For instance, the presence of a set of variables *a\_w1*, *a\_w2*, ..., *a\_w10* means that there are up to 10 values for the variable *a* (e. g. the item corresponding to variable *a* was measured repeatedly in a questionnaire loop) relating to a row entity (e. g. a person or a school episode). Also, wide-format suffixes indicate different versions of instrument IDs. Of course, there are cases where suffix *\_w#* directly relates to wave-specific values of the underlying variable.

### ■ Confidentiality suffix

This suffix pertains to all variables that were anonymized (see section 6 on page 15). The suffix indicates a variable's degree of anonymization. This suffix may either stand alone (e. g., country of birth: *t80a\_R*) or be combined with other suffixes (e. g., coded mother tongue: *t83\_g2R*).

**O** on site; data on this variable are only available on site

**R** remote access; data on this variable are available on site or via RemoteNEPS

**D** download; data on this variable are available via all three modes of access

Element	Definition
BW	<b>Indicator of "G8 Reform Study in Baden-Wuerttemberg"</b>
[filename]	<p><b>Filename conventions</b></p> <p><i>Prefix:</i> x = pooled cross-sectional file</p> <p><i>Keyword/mnemonic:</i> A keyword or mnemonic indicates the content of the corresponding file (e.g., xInstitution contains data from the subject-leaders and school heads)</p> <p>Filenames of generated datasets do not have a prefix and always start with a capital letter (e.g., Profile)</p>
[D,R,O]	<p><b>Confidentiality Level</b></p> <p>D = Download version</p> <p>R = Remote access version</p> <p>O = Onsite version</p>
[#]-[#]-[#](_beta)	<p><b>Version</b></p> <p><i>First digit:</i> denotes the main release number; since the data from the G8 Reform Study in Baden-Wuerttemberg had been released for two cross-sections at once and integrated into the same file in long format, the main release number has only changed once, as data from the third and last wave has been released in March 2015.</p> <p><i>Second digit:</i> indicates major updates; major updates affect the data structure (e.g., release of imputed datasets); updating your syntax files may be necessary.</p> <p><i>Third digit:</i> indicates minor updates; minor updates affect the content of cells but not the data structure; updating your syntax files is not necessary.</p> <p>_beta: this suffix indicates a preliminary release which allows users to test the data in advance of the main release. The beta version is no longer available after the main release.</p>

Table 1: Naming conventions of file names

## 2.3 Missing values

### 2.2.1 Special conventions for variables in test data

Naming of variables corresponding to test items (found in datafile `xTargetCompetencies`) follow an alternative nomenclature. Variable names usually start with the test subject (one to three characters, e. g. `ma` for Math) and are suffixed by `_c`, indicating a scored item. Additional variables like sum scores (e. g. `dgc_i_1`) or weighted maximum likelihood estimates (WLE, e. g., `ma2_sc1`) can be identified by a variable name diverging from this pattern.

### 2.3 Missing values

We provide different missing codes for different situations of missing values. In general, we distinguish between missing codes indicating sorts of item nonresponse, not applicable missings, and edition missings. When working with the NEPS data make sure that you correctly process those codes in your statistical package. Most packages available provide functions for defining missing values. Table 2 on the next page provides an overview of missing codes you will encounter in the NEPS data.

[htp]

We distinguish between three types of missing values:

- *Item nonresponse* occurs if a person did not respond to a question.
  - The most common instances of item nonresponse are refusals (-97) and don't knows (-98).
  - Implausible values are coded by a -95 missing value.
  - For competence data there is a special missing code -94 that indicates that a test item has not been reached, because the target person quit the test somewhere before this item.
  - further missing codes (-5/-6/-20, ..., -29) pertain to specific nonresponse categories (e. g., -22 is used as a missing code in the grade records if the student was not enrolled in the specific subject).
- *Not applicable* denotes missing data that occur because the item does not apply to a person. This category comprises three kinds of missings.
  - The first concerns samples: If a question is not included in a sample-specific questionnaire, the code -54 is assigned to all respondents from this sample. This code is used also for the more general case where values of a variable are not available due to design issues.
  - The second concerns individuals: If a question does not apply to a person, it is coded "not applicable" either by the respondent's or the interviewer's remark (-93) or like it is the case for computer-assisted interviews automatically by the survey instrument ( . = filtered).
  - Missings that occur for unknown reasons are coded by -90; this especially happens in PAPI questionnaires, where the cause for a respondent not answering a question cannot be determined.
- *Edition missings* are defined in the process of data editing.
  - Implausible values are recoded into missing (-52).
  - Sensitive information which is only available via RemoteNEPS and/or on site access is anonymized (-53).

Code	Missing
<b>Item nonresponse</b>	
-97	refused
-98	don't know
-94	not reached (only applicable for competence tests)
-5/-6/-20,...,-29	item-specific missing with informative value labels (also, see additional missings below)
<b>Not applicable</b>	
-54	missing by design (mostly: not included in sample-specific instrument of this wave)
-90	unknown missing
-93	does not apply
-96	not in list
-99	filtered (in PAPI mode)
.	filtered / system missing (in CATI/CAPI mode)
<b>Edition missings (recoded into missing)</b>	
-52	implausible value removed
-53	anonymized
-55	not determinable
-56	not participated
<b>Additional missings (particularly for grades and course)</b>	
-22	not enrolled
-23	not admitted
-24	not passed
-25	not valid
-27	not present

Table 2: Overview of missing codes

- Coding schemes are used to generate variables (e. g., occupational coding). If the information from the original data is not sufficient to generate a value, we assign the missing code “not determinable” (-55).
- If a person was not present during the interview, or did not fill out a questionnaire although it was administered to him or her, the concerning variables are assigned the missing code “not participated” (-56). This missing code is special in so far as target persons lacking interview data (e. g., due to illness) usually are not entailed in the corresponding datasets. In the special case of one dataset integrating multiple waves widely this missing code is assigned.

### nepsmiss: Recoding missing values in Stata

We offer a Stata ado file on our web portal which automatically recodes all missing values into extended missing values (.a, .b, etc.), and vice versa, while preserving value labels. We generally recommend running `nepsmiss` before any further data preparation. See section 8 for further information on how to install and update the `nepsmiss` command.

## 3 Sampling and surveying procedures

### 3.1 Overview

Many German federal states are currently reforming the curriculum and the organization of secondary schools (*Gymnasiale Grund- und Oberstufe*). In general, these changes target a stronger emphasis on general education, a restriction of the differentiation in the *Leistungskurs-Grundkurs-System* during the last two years in the *Gymnasiale Oberstufe*, the switch from input-oriented towards output-oriented measurement of educational achievement as well as the overall duration of secondary education. The NEPS-G8 Reform Study in Baden-Wuerttemberg aims to study the possible effects of such a reform – the reduction of the duration from nine to eight years in *Gymnasiale Grund- und Oberstufe*.

Three cross-sectional surveys are conducted for the graduation years

- 2011: last year with G9 only; **wave 1** (NEPS study A72)
- 2012: G8 and G9 combined (*Doppeljahrgang*); **wave 2** (NEPS study A73) and
- 2013: G8 only; **wave 3** (NEPS study A74)

The target population of the study comprises all grade 12/13 students in 2011, 2012 and 2013 in Baden-Wuerttemberg.

The students participated in a questionnaire, achievement tests (*Fachleistungstests*) in the fields of Mathematics, Physics, Biology, and English, and a test on cognitive abilities. In addition, relevant context persons were surveyed via questionnaire – that is school principals and those teachers coordinating German, English and Maths courses – the heads or leaders of the departments. Field work was conducted by IEA DPC (IEA Data Processing and Research Center, Hamburg).

population	wave 1	wave 2	wave 3	pooled
Students (N-sampled initially)	1858	3826	1909	7593
Students general participation (N with consent)	1341	2577	1292	5210
Students participated in questionnaire (PAPI)	1281	2392	1211	4884
Students participated in achievement tests	1284	2395	1214	4893
Students attained reduced duration (G8)	52	1284	1179	2515
Subjectleader English (PAPI)	46	44	39	129
Subjectleader German (PAPI)	42	39	41	122
Subjectleader Math (PAPI)	42	44	37	123
Head of School (PAPI)	40	43	44	127
Grading information available	1341	2542	1292	5175

Table 3: Overview of samples and survey instrument participation

### 3.2 Sampling and response rates

The stratified sample consists of 1858 (wave 1), 3826 (wave 2) and 1909 (wave 3) randomly drawn grade 12/13 students from 48 randomly selected grammar schools in Baden-Wuerttemberg. See table 3 for a detailed overview of participation.

For further details on the sampling and weighting procedure see the technical report on weighting (Schönberger et al., 2015).

### 3.3 Competence testing and students questionnaire

All achievement tests in the fields of Mathematics, Physics, Biology, and English, and the test on cognitive abilities as well as the students questionnaire (PAPI) were conducted in the schools within one day in mid May 2011 (wave 1), at the end of April/early May 2012 (wave 2) and mid May/early June 2013 (wave 3).

In the first test session in the morning the students had to perform either in one of four different versions of a Physics test (indicated by variable `tx80211_w5` in datafile `Profile`) or in one of four different Biology test versions (see variable `tx80211_w2`). This first block was intended to last 45 minutes. The second test session was dedicated to the domain of English (instrument ID given in `tx80211_w4`) or Math (`tx80211_w6`) and was administered in 30 minutes.

After that – and a break – the students were asked to participate in

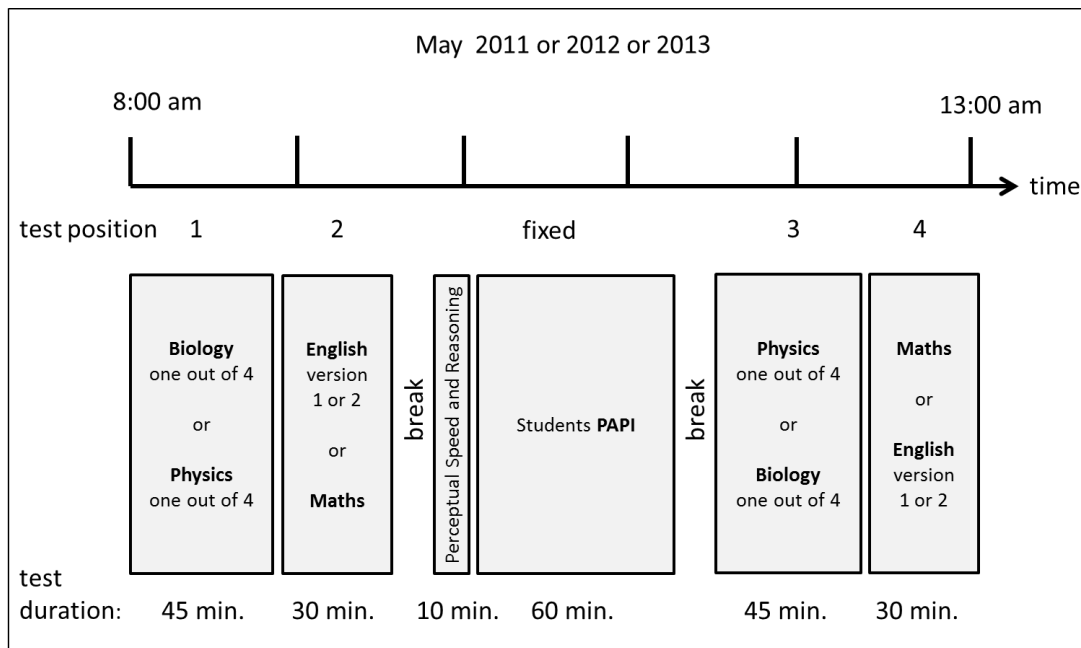
- the test on cognitive abilities: perceptual speed (duration 3\*30 seconds) and reasoning (duration 3\*3 minutes); overall about 10 minutes and
- the students questionnaire; duration about 60 minutes.

After a second break, the third and fourth achievement tests were administered in the same manner as the first ones, starting with either a Biology or Physics version and ending with Math or English forms again.

To identify the relative position of a subject-specific test, separate order variables are provided in file `Profile` (`orderbio`, `orderen`, `orderphy` and `ordermat`). For a detailed description of the competence tests find additional information on our website at

→ [www.neps-data.de](http://www.neps-data.de) > Research Data > Additional Study Baden-Wuerttemberg

### 3.4 School heads and subject heads questionnaire



**Figure 1:** General procedure on students' competence/ability tests and surveying in school

The students questionnaire consists of central socio-demographic questions (age, sex, country of origin, mother tongue etc.), questions on interests, aspirations, leisure activities, health, life satisfaction, the familial background from the students perspective, class climate, school subject specific evaluations as well as individual tuition and learning traits and finally on questions regarding reform aspects and consequences. Overall, the whole testing and survey procedure at the schools took around five hours.

### 3.4 School heads and subject heads questionnaire

All School heads and subject heads who are responsible for the subjects in German, English or Maths in the 12/13 grade were asked for *subject specific* evaluations – also in respect to the reform and finally for some school- and person specific background information. Note however, that the questionnaire (PAPI) is subject specific and not a unique teacher instrument. Hence, teachers as subject heads may have provided information to more than one questionnaire, if they preside several subjects at once.

### 3.5 School grades and tracking information

School grades from German, Maths, English, Physics, Biology and Chemistry in all four terms of the senior years of all 12/13th grade students of the participating schools as well as the grading in the five subjects of the final exams and the final grade point average were collected. This was done retrospectively by the IEA DPC at the end of the 12/13th grade in 2011, 2012 and 2013 via the school coordinators<sup>3</sup>, the school principals and the school's databases. The grades and results of those students not participating in the study had been sent and processed

<sup>3</sup> local teachers or principals responsible for organizing all activities within schools and classes that are necessary to realize the school survey

in a completely anonymized and aggregated form and were used for the calculation of weights by the method group of the NEPS only. See technical report on weighting (Schönberger et al., 2015) for details.

Additionally, the students' sex, course type, course participation, instrument version and/or rotation, and legal age was collected during the tracking process via the school coordinators. Similar tracking data on all subject heads were also collected and sent back all together to the DPC as anonymized lists.

## 4 Datafiles

As introduced above, the NEPS G8 Reform Study in Baden-Wuerttemberg collects data of different types and from different sets of respondents: student data (paper questionnaire, competence and ability tests), subject heads data (take home paper questionnaire), school heads data (take home paper questionnaire), tracking data from school coordinators, and, last not least, students' grades and exam results provided by schools.

Grading data and some tracking information are mainly integrated into the so called `Profile` dataset, while students' data resembled into a separate dataset `xTarget`. The information given at school level by subject heads and school heads are integrated into an institutional data file `xInstitution` in wide format. Competence test data of all domains and types are finally merged into another separate file: `xTargetCompetencies`.

In order to provide a most convenient data structure, the data from all three separate cross sections in 2010/2011, 2011/2012 and 2012/2013 are already pooled within the respective files. All datasets contain an indicator variable `wave`, that identifies the pre/post-reform data, and can easily be merged to all other datasets via `ID_t` and, in case of `xInstitution`, `ID_i`. Using `Profile` as the primary merging unit is recommended (see figure 2 on page 13).

Remember, however, that `wave` is not a panel wave indicator like it is in the datafiles of the starting cohorts of NEPS. Since the G8 Reform Study in Baden-Wuerttemberg asked each student only once<sup>4</sup> it is a cross-section indicator!

### 4.1 Pooled cross sectional target file: `xTarget`

The file `xTarget` contains all the data from the students' questionnaire as well as coded educational and occupational aspirations and parents occupations (suffix `_g1,...,_g16`).

### 4.2 Pooled cross sectional competencies file: `xTargetCompetencies`

This file contains scored and scaled data from the competence tests in Math, Physics, English and Biology, as well as the test data on general cognitive abilities. Instrument IDs (`tx80211_w1, ..., tx80211_w6`) indicating rotation or version type and variables indicating the relative position of the four competence tests during test day (`orderbio, ..., ordermat`) from the tracking

<sup>4</sup> However, the schools are sampled only once for all waves, which might lead to the situation that perhaps the same subject heads and very likely the same school heads are asked in more than one wave – but for different situations in grade 12/13. The design of the study does not allow the reidentification of teachers in wave two or three, so there is no way to deal with this special type of clustering.



## 4.3 Pooled cross sectional file for teachers and heads: xInstitution

lists can be found in `Profile` and have to be merged to `xTargetCompetencies` if needed. The general participation and design specific missing structure in each domain specific test can be derived from the corresponding missing information coding of the instrument IDs (-56 and -54). Also, see figure 1 on page 10 for a quick overview.

### 4.3 Pooled cross sectional file for teachers and heads: xInstitution

This file integrates the subject leader (German, English and Maths) and school-heads paper questionnaire responses for all cross sections. That is, each case in the datafile (identified uniquely by variables `ID_i` and `wave`) combines information from all participating teachers from this school.

### 4.4 Linking and method file: Profile

To facilitate usage and enable a quick overview on the different types of data and cases available, we provide the dataset `Profile`. This dataset contains the case number (`ID_t`) of all target persons (the students) who participated at least in one instrument. Therefore it is also recommended as a starting point for merges between different respondents (see figure 2 on the next page). Additionally, the following central information is provided:

- `ID_t`: unique student identifier
- `ID_i`: unique school identifier<sup>5</sup>
- `ID_tg`: unique testgroup identifier
- `wave`: indicating the three separate cross sections for school year 2010/2011 (`wave=1`), 2011/2012 (`wave=2`) and 2012/2013 (`wave=3`)
- `txg8`: whether a student graduated school in a reduced duration of eight years (G8, `txg8=1`) or still attended school for nine years (G9, `txg8=0`) can be distinguished by this variable. Please note that some students already attained a G8 cycle voluntarily in 2010/2011 (`wave=1`; also see table 3 on page 9).
- `tx80211_w1-6`, `ex80211_w1-3`, `hx80211`: variables indicating whether data from a specific type of instrument, mode, or respondent is available have not been provided separately, but this information can be derived from the respective missing codes of the instrument ID variables
- `orderbio`, `orderen`, `orderphy`, `ordermat`: variables describing the sequence of tests (also see section 3.3 on page 9)
- Sex as well as year and month of birth from the tracking information
- `ID_cger`,...`ID_cch`: cluster information for the attained class can be derived from these
- Additionally, the grade point information (*Kurspunkte*) from the school data bases for subjects of Maths, German, English, Biology, Physics and Chemistry and exam results for five relevant exam-subjects (German, Maths, foreign language, additional Major (*Kernfach*) and the oral examination subject). For example, `grdmat_w1` provides grade points in Math course from the first half year in grade 11.

<sup>5</sup> Unique only within wave!

Analytical	Wave (wave) Requirement level (e.g. hrsbio: weekly hours biology) Individual (ID_t)
Institutional	School (ID_i) Subject (e.g., crsna_w4: 4. Abiturfach) Requirement level (e.g. hrsbio: weekly hours biology) Specific course (e.g., ID_cmat) Individual (ID_t)
Procedural	Wave (wave) School (ID_i) Rotation order/position (e.g., orderphy) Instrument version (e.g., tx80211_w5)

Table 4: Different general perspectives and possible levels/clusters within the data

#### 4.5 Clustering and merging within a multilevel data structure

The data structure resulting from this study has a lower multilevel complexity for school studies. Although there are many different possibilities to construct levels or clusters within the data, some emerge directly from the analytical, institutional and procedural perspective and should be of special interest. Table 4 gives a short overview on central perspectives and the main levels of interest. Of course the same level might be of interest for more than one perspective.

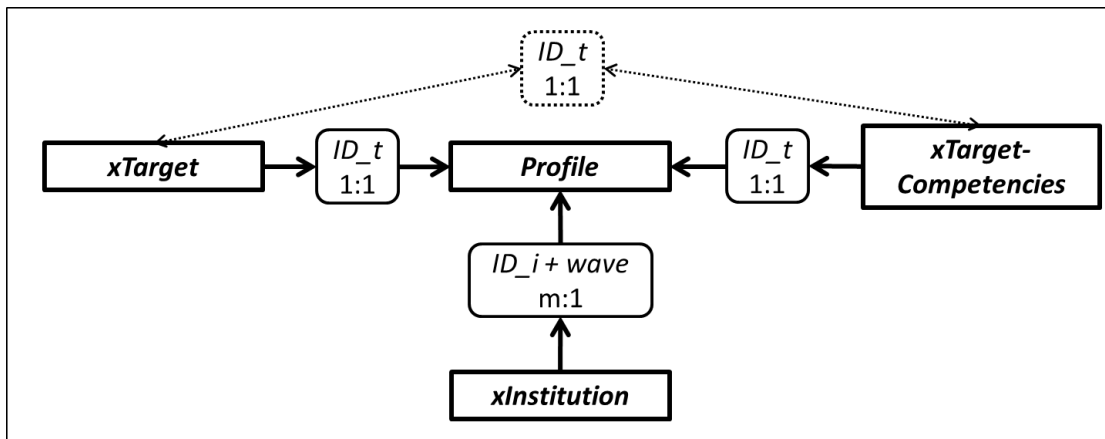


Figure 2: Recommended merging relation for the datafiles

The most simple and probably most common type of merge is between the Profile data and the xTarget or xTargetCompetencies files. Using ID\_t as a unique identifier and a one to one merging command is enough to perform the merge. To keep all cases in this step, it is always recommended to merge against Profile first (see section 7 on page 18 for Stata and SPSS examples).

Linking the data from the xInsitution file with the other datasets is easily done via a m:1 merge using variables ID\_i and wave.

## Note

The NEPS invested a lot to ensure the integrity of these data. However, we strongly recommend you to examine the data critically when you work with this release. Furthermore, you should always consult the questionnaire/s to obtain a precise understanding of how the data have been collected.

# 5 Generated variables and weights

## 5.1 Coding

All string variables based on occupations, vocational training information and subjects of study of the respondents and their parents have been coded. Table 5 presents an overview on these coded variables; variables that are derived from those are suffixed by an additional `_g#`, which are particularly useful if you are interested in cross-national comparisons. See section 2.2 on page 3 for their meaning.

Variable	Description	German Description
t63_gX	likely vocational training	wahrscheinliche berufliche Ausbildung
t64_gX	likely full-time school education program	wahrscheinliche vollzeitschulische Ausbildung
t65_gX	likely professional activity	wahrscheinliche berufliche Tätigkeit
t75a_gX	Mother: occupation	Mutter: Beruf
t75b_gX	Father: occupation	Vater: Beruf

Table 5: Overview for coded variables (German labels also provided)

Tables that show, how BLK (classification of occupations according to Blossfeld, cf. Blossfeld 1985; Schimpl-Neimanns 2003), the ISCED-97 (UNESCO 2006) and CASMIN (Lüttinger & König 1988) classes are coded, are presented in section A.1 on page 26.

## 5.2 Weights

The weighting variables provided in the `Profile` dataset are indicated by the prefix `w_`. We provide sampling weights like (`w_*`) and standardized versions (`w_*_std`) for the participating schools as well as the students. There are separate weighting variables for students participating in the different waves as well as G8 and G9 students in wave two. Also, detailed information on the construction of weights and how to use them will be available in the technical report on weighting (Schönberger et al., 2015). For a more general discussion on the usage of sampling weights for model estimation see Rohwer (2011).

## 6 Anonymization

This section gives an exhaustive explanation of all disclosure risk minimisation techniques applied before dissemination of the data. For a quick reference what is done to the datasets in detail and on which level you will find your desired information, please see section A.2 on page 29, where all affected variables are listed.

### 6.1 Specifications

To ensure the best possible confidentiality protection of individuals and individual micro data, the National Educational Panel Study complies with strict international standards. Operationalise those, they have been abstracted to the following two criteria:

1. the disseminated data has been transferred to so called *de facto anonymous data*. Identifiable information is coarsened or cut off and kept securely to minimise the risk of statistical disclosure.
2. the use of data is strictly confidential and for statistical purposes only. The closed contract only grants access to members of the scientific community. This contract has a vast amount of legal stipulations, one of them being a large fine which applies for the realisation of re-identification on purpose. Therefore, the disseminated data is highly protected by law and allows a more flexible range of available data.

To pick up the latter, the NEPS has made a huge effort regarding legal regulations to offer as much analysis power of data as possible. This *paradigm of information esteem* reveals the fact that conducted measures of statistical disclosure control are few. Also, if there really was a need for modification, only non-perturbative methods were used.

### 6.2 Onion-shaped model

The NEPS grants the user three different modes of data access: (1) **OnSite**, which stands for the opportunity to use the secured infrastructure made available at the NEPS in Bamberg, (2) **RemoteNEPS**, which is a progressive remote access technology providing a virtual desktop, and finally (3) **Download**, indicating the possibility to fetch data via a secure web portal.

These given access modes have been originated to allow anonymization routines for a subtle differentiation of information. The three resulting levels of anonymization define as follows:

- data provided **OnSite** is generally not further anonymized. However, even those data has been rendered *de facto anonymous*, for no disclosure risk to persist. All information contained remains completely sane. Although users have to deal with limited possibilities of data access (i.e. supervised import and export of their results), they are free to work with all data available at the NEPS in a secure environment.
- access via **RemoteNEPS** is considered equivalent to **OnSite**, hence most of the data stays complete.

## 6.2 Onion-shaped model

- as *Download* is assumed to be the most hazardous access mode<sup>6</sup>, some more anonymization techniques are applied to the dataset.

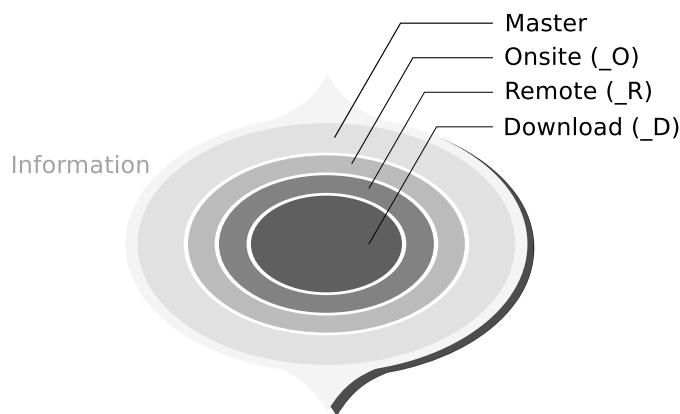
Obviously this approach results in three different versions of all involved datasets. To enable a consistent structure, these data files always contain the entire set of variables; it is their content which differs through the three levels.

As normally there is no need to resign aggregated variables in the higher levels (i.e. *OnSite* or *RemoteNEPS*), those are already defined as a surplus to the original variable in the *OnSite*-version. Stepping down to *RemoteNEPS* the content of related variables too sensitive for this level is overwritten with an exclusive missing code – an operation which we define as *purging*. Note that system missing values are not affected, allowing the user to differ between value existence and nonexistence. This still is a valuable additional information. Same applies to *Download*.

While there is no explicit documentation to this fact, it should remain clear that this procedure accumulates, i.e. purged content under *RemoteNEPS* is therefore neither included in *RemoteNEPS* nor in *Download*.

This *onion-shaped* model provides both, ease of (1) use of different sensitivity models (e.g. preparing an analysis using the *Download* dataset and conducting it afterwards using the *OnSite*-data) and (2) documentation, for the subject of documentation is the most sensitive level (*OnSite*), with *RemoteNEPS* and *Download* levels being a subset of these data.

The fourth layer *master* depicted below contains every material which is needed during data processing by the NEPS, but is not meant for the scientific community to be usable.



**Figure 3:** Onion-shaped model defining the different anonymization levels

Formal, this model results in a single letter suffixed to dataset and variable names. All datasets available *OnSite* only are marked with an additional *\_O*, those available via *RemoteNEPS* with *\_R* and *Download* files with *\_D*. The same procedure applies when it comes to variable differentiation. A variable which is only available *OnSite* has been suffixed with *\_O*. In *RemoteNEPS* access or *Download*, this variable is still present but purged. If there is an alternate version (mainly with coarsened content) for *RemoteNEPS* (suffix *\_R*) or *Download* (suffix *\_D*), those can be used. As said before, these are already integrated in the *OnSite* version.

<sup>6</sup> 'hazardous' in terms of: the downloaded content is no longer under physical control of the NEPS

Topic	<i>OnSite</i>	<i>RemoteNEPS</i>	<i>Download</i>
International <sup>7</sup>	full data	full data	collapsed
String variables	anonymised	n/a	n/a
Institutional	full data	full data	n/a

Table 6: Availability of sensitive data

### 6.3 Conducted measures

Keeping the usability and the paradigm of information esteem in mind, only very few alterations are actually done to the dataset. These modifications always account for the fact that information may never be lost completely, but aggregated into coarse categories or variables. Please note that all information is still available somewhere and that only *RemoteNEPS* and (mainly) the *Download* version are constraint in this matter. In fact, roughly 40 variables are modified in some way – which is about three percent of the whole dataset volume.

Please refer to section A.2 on page 29 for a complete overview of all variables which fell victim to anonymization.

The following gives an explanatory overview of all measures conducted.

**countries and languages** Complete information corresponding to (international) localization, nationality or languages is only available *OnSite* or via *RemoteNEPS*. Variables comprised in the *Download* Scientific Use File (SUF) are aggregated into German and Non-German.

**open ended strings** All string variables containing actual text are purged in the *RemoteNEPS* version. The information remains accessible *OnSite*. However, all text entries have been reviewed by staff to ensure that absolutely no re-identificational material is included.

**institutions** For studies conducted in an institutional context, special focus of anonymization has been directed to protection of institutional data, i.e. information about kindergarten and schools, but also educators and teachers. This includes the complete datafile *xInstitution*, but also basic structural details about the kindergarten group or school class. Furthermore, personal information about educators and teachers is treated more securely. You will find detailed information about these subjects from *RemoteNEPS* onwards.

<sup>7</sup> international geographical information (e.g., nation states, national languages)

# 7 Examples

This section gives some examples of how to merge different datasets. We provide you with the code to run the examples in Stata and SPSS.

## 7.1 Example 1—Merging data from xInstitution and xTarget via Profile

In the example shown below we merge data on the respondent's usage of additional private lessons in English (t46d in xTarget) and the English subject leaders evaluation on impact of the reform on the competence level in English (e12a in xInstitution). Note that you will need *Remote-Access* (see section 1.3 on page 2) to execute this syntax, as xInstitution is only available there or onsite, of course.

### Stata example 1: Merging information from xInstitution and xTarget data via Profile

```
/******  
Example 1: Merge specific information from Institution to xTarget via Profile  
  
Procedure:  
1. Open Profile  
2. Merge variables from xInstitution to Profile with a m:1-merge by ID_i and wave  
3. Keep the relevant variables  
4. Merge variables from xTarget to Profile with a 1:1-merge  
  
(note: replace ${version} to your file version)  
*****/  
  
global version R_3-0-0  
  
use "BW_Profile_${version}.dta", clear  
  
merge m:1 ID_i wave using "BW_xInstitution_${version}.dta", keepusing (e12a) nogen  
  
keep ID_t wave e12a  
  
merge 1:1 ID_t using "BW_xTarget_${version}.dta", keepusing (t46d) nogen  
  
* recode missing values  
nepsmis*  
  
* calculate a correlation  
pworth e12a t46d, sig
```

**SPSS example 1: Merging information from xInstitution and xTarget data via Profile**

```

*****
* Example 1: Merge specific information from Institution to xTarget via Profile
*
* Procedure:
* 1. Open Profile
* 2. Merge variables from xInstitution to Profile with a m:1-merge by ID_i and wave
* 3. Keep the relevant variables
* 4. Merge variables from xTarget to Profile with a 1:1-merge
*****.

* load CohortProfile.
GET FILE='BW_Profile_R_3-0-0.sav'.
DATASET NAME Profile.
SORT CASES BY ID_i.
EXECUTE.

GET FILE='BW_xInstitution_R_3-0-0.sav'.
DATASET NAME xInstitution.
SORT CASES BY ID_i.
EXECUTE.

MATCH FILES
  /FILE = 'Profile'
  /TABLE = 'xInstitution'
  /BY ID_i wave
  /KEEP = ID_t wave e12a.
SORT CASES BY ID_t.
DATASET NAME Profile_xInstitution.
EXECUTE.

DATASET CLOSE Profile.
DATASET CLOSE xInstitution.

GET FILE='BW_xTarget_R_3-0-0.sav'
  /KEEP = ID_t t46d.
DATASET NAME xTarget.
SORT CASES BY ID_t.

MATCH FILES
  /FILE = 'Profile_xInstitution'
  /FILE = 'xTarget'
  /BY ID_t.
DATASET NAME Profile_xInstitution_xTarget.
EXECUTE.

DATASET CLOSE Profile_xInstitution.
DATASET CLOSE xTarget.

* calculate a correlation.
CORRELATIONS VARIABLES = e12a t46d.

```



## 7.2 Example 2—Using Profile

### 7.2 Example 2—Using Profile

In the next example, we try to combine information from Profile with content from xTarget and analyse the satisfaction with the school and the sex of child.

#### Stata example 2: Using Profile

```

/*****
Example 2: Using Profile

Procedure:
1. Open Profile
2. Merge variables from xTarget to Profile with a 1:1-merge by ID_t
3. Analyse Satisfaction School and Sex of Child

(note: replace ${version} to your file version)
*****/

global version D_3-0-0

* load Profile
use "BW_Profile_${version}.dta", clear

* merge variable satisfaction: school from xTarget
merge 1:1 ID_t using "BW_xTarget_${version}.dta", keep(master match) keepusing (t1f)
nogenerate
keep ID_t wave t1f tx80501

* for all variables recode NEPS missings to Stata missings
nepsmis _all

* Check status of wave participation
fre wave

* Analyse: Satisfaction School and Sex of Child
* two-way tables of frequencies
tabulate t1f tx80501, column

* calculate means
mean t1f, over (tx80501)

```

### SPSS example 2: Using Profile

```
*****
* Example 2: Merge specific information from xTarget to xProfile
* Procedure:
* 1. Open xProfile
* 2. Merge variables from xTarget to xProfile with a 1:1-merge by ID_t
* 3. Analyse Satisfaction School and Sex of Child
*****

* load Profile.
GET FILE='BW_Profile_D_3-0-0.sav'.
DATASET NAME Profile.
SORT CASES BY ID_t.
EXECUTE.

GET FILE='BW_xTarget_D_3-0-0.sav'.
DATASET NAME xTarget.
SORT CASES BY ID_t.
EXECUTE.

MATCH FILES
  /FILE = 'Profile'
  /TABLE = 'xTarget'
  /BY ID_t
  /KEEP ID_t wave tx80501 t1f.
DATASET NAME Profile_xTarget.
EXECUTE.

DATASET CLOSE Profile.
DATASET CLOSE xTarget.

* check status of wave participation.
FREQUENCIES wave.

* Crosstab analysis: Satisfaction School and Sex of Child.
CROSSTABS t1f BY tx80501
  /CELLS = COUNT COLUMN.

* calculate means.
MEANS t1f BY tx80501.
```

## 8 Tools for Stata users

Thanks to Stata's quite versatile data format, NEPS Scientific Use Files (SUFs) provide additional information with easy-to-use tools. LIfBi's Research Data Center (RDC) provides a package of additional Stata programs ("ado files"), in order to present this additional information to the user as clearly as possible.

### 8.1 Multi-lingual data sets

The Stata files from the RDC offer multi-lingual variable labels and value labels (currently: German and English). Stata users can easily switch between these languages using Stata's label language command:

```
label language // ... to display saved label languages
```

```
Language for variable and value labels
```

```
Available languages:
```

```
de  
en
```

```
Currently set is: . label language de
```

```
To select different language: . label language <name>  
(more output omitted)
```

```
label language en // ... to change label language to -en-
```

### 8.2 Data signatures

Upon publishing, a signature of each and every data set is generated and accordingly saved. Stata users can use Stata's `datasignature confirm` command to check if the data set has been modified since its publication:

```
datasignature confirm
```

```
(data unchanged since 28jun2014 21:41)
```

### 8.3 NEPStools

As mentioned before, the LIfBi RDC provides Stata programs as Stata packages.

### 8.3.1 Installation and updating

The package can be easily installed through Stata's built-in installation mechanism:

```
net install nepstools, from(http://nocrypt.neps-data.de/stata)
```

```
checking nepstools consistency and verifying not already installed...
installing into ...
installation complete.
```

Updates should be carried out regularly, since the program package is updated regularly as well:

```
adoupdate nepstools , update
```

In case you do not have access to Stata's regular package installation mechanism, feel free to contact the LfBi RDC to obtain the package as a ZIP file. It is highly recommended to first read the online documentation at → [www.neps-data.de > Data Center > Research Data > Stata tools](http://www.neps-data.de/Data_Center/Research_Data/Stata_tools) as well as the Stata help files for each of the provided programs (`help <program>`) carefully.

### 8.3.2 *nepsmiss*—Recoding missing values

This program automatically recodes all of the numeric missing values from the NEPS SUFs (-97, -98, etc.) into Stata's "Extended Missings" (.a, .b, etc.). In contrast to Stata's built-in commands `mvencode` and `mvdecode`, value labels are correctly recoded and a standard list of missing values has already been predefined. An illustration is printed in table 7.

```
nepsmiss t731454
```

This process can be reverted with the option `reverse`:

```
nepsmiss t731454 , reverse
```

We generally recommend using `nepsmiss` to recode all of the missing values after any data preparation and prior to an analysis of the working data set (`nepsmiss _all`).

ID_t	wave	t731454		ID_t	wave	t731454
8010851	1	4		8010851	1	4
8010851	2	-97	⇒	8010851	2	.b
8012254	1	-54	⇒	8012254	1	.c
8012254	2	5		8012254	2	5
8002388	1	1		8002388	1	1
8002388	2	-98	⇒	8002388	2	.a

Table 7: Data set comparison before and after using *nepsmiss*

## 8.3 NEPStools

### 8.3.3 *infoquery*—Display additional metadata

As you may already know, additional metadata is saved to NEPS SUFs. The metadata is saved as characteristics in the Stata data sets. The program *infoquery* is meant to display these attributes directly in Stata's dialog mode:

```
label language en
infoquery t514001
```

```
variable t514001:
-----
NEPS_instname      zufried1
NEPS_sufname       t514001
NEPS_questionnumber 20110a
NEPS_outputfilter_en goto 20110b
NEPS_questiontext_en To start, I'd like to ask you a few questions about your
                    current level of satisfaction with different aspects
                    of your life. Please answer each question on a scale
                    of 0 to 10. "0 means that you are "completely"
                    unsatisfied, while a "10 means that you are
                    "completely "satisfied.
                    You can indicate the degree of satisfaction by
                    choosing a number between 0 and 10.
                    How satisfied are you with your life, all in all?
```

### 8.3.4 *charren*—Rename variables to survey names

As part of the data preparation for the NEPS SUFs, the used variable names from each of the survey instruments are renamed as SUF variable names. However, some users, especially those who have become accustomed to the original names, may want to reverse this data preparation step. The program *charren* easily allows switching between these versions, by using the saved metadata in the Stata data sets.

```
charren t514001 , to(NEPS_instname) verbose
```

```
Info: will rename t514001 to zufried1
```

Note that a reverse-search is also possible: Even if the user only knows one instrument name, the program will reliably find the respective SUF variable and rename it correctly.

```
charren zufried1 , to(NEPS_instname) verbose
```

```
Info: zufried1 is not a variable name in current dataset; searching for
zufried1 in specified search space
Info: will rename t514001 to zufried1
```

# 9 Further information

Please visit our web portal for further information and comprehensive documentation resources such as PAPI questionnaires, how-to guides, technical reports, and the codebook.

→ [www.neps-data.de](http://www.neps-data.de) > Data Center > Research Data > Additional Study Baden-Wuerttemberg

For further support, please contact the LifBi Research Data Center:

E-mail: [fdz@lifbi.de](mailto:fdz@lifbi.de)

Web: → [www.neps-data.de](http://www.neps-data.de) > Data Center > Contact the Data Center

Phone: +49 951 8633511 (Mo-Fr 10:00-12:00 and 14:00-16:00)

### Participation in the NEPS user trainings

Furthermore, the Research Data Center offers training courses on a regular basis. These courses introduce the research design of the NEPS, the structure of datasets, terms and conditions of data usage, issues of privacy and data protection, and so on. A central module of the courses consists of hands-on work with the NEPS data supervised by our staff. As skill levels, research interests, and methods vary greatly across users and disciplines, we will offer a comprehensive portfolio of seminars ranging from introductory topics on a rather general level to advanced methodological courses.

# A Appendix

## A.1 Occupational coding

**Table 8:** Coding of BLK from (Blossfeld, 1985)

Key	English	German
1	[AGR] Agricultural occupations	Agrarberufe
2	[EMB] Common manual occupations	Einfache manuelle Berufe
3	[QMB] Skilled manual occupations	Qualifizierte manuelle Berufe
4	[TEC] Technician	Techniker
5	[ING] Engineer	Ingenieure
6	[EDI] Common services	Einfache Dienste
7	[QDI] Skilled services	Qualifizierte Dienste
8	[SEMI] Semiprofessions	Semiprofessionen
9	[PROF] Professions	Professionen
10	[EVB] Common commercial and administrative occupations	Einfache kaufmännische und Verwaltungsberufe
11	[QVB] Skilled commercial and administrative occupations	Qualifizierte kaufmännische und Verwaltungsberufe
12	[MAN] Manager	Manager

**Table 9:** Coding of ISCED-97 (United Nations Educational, Scientific and Cultural Organization (UNESCO), 1997)

Key	English	German
0	0A/1A Inadequately completed general education	kein Abschluss
1	2B Lower general education	Haupt-, Volksschulabschluss, Berufsvorbereitende Maßnahme
2	2A Intermediate general education	Mittlere Reife, Realschulabschluss
3	3A Full maturity certificates (e.g., the Abitur, A-levels)	Fachhochschulreife, Hochschulreife
4	3B Basic vocational training, Vocational full time school, Health sector school (less than two years), civil servant of the lower grade, vocational basic skills	Lehre, Berufsfachschule, Fachschule des Gesundheitswesens (weniger als zwei Jahre), Beamter einfacher Dienst, berufliche Grundkenntnisse
5	3C Civil servants of the medium grade	Beamter mittlerer Dienst

(...)

**Table 9:** (continued)

Key	English	German
6 4A	Full maturity certificates (e.g., the Abitur, A-levels) (second cycle)	Fachhochschulreife, Hochschulreife (second cycle)
7 4B	Basic vocational training, Vocational full time school, Health sector school (less than two years), civil servant of the lower grade, vocational basic skills (second cycle)	Lehre, Berufsfachschule, Fachschule des Gesundheitswesens (weniger als zwei Jahre), Beamter einfacher Dienst, berufliche Grundkenntnisse (second cycle)
8 5B	Diploma (vocational and other specialised academies, College of public administration), Qualification of a two or three year Health-Sector School, Master's/technician's qualification	Fach- und Berufsakademische Abschluss, Verwaltungsfachhochschule, Fachschule des Gesundheitswesens (mindestens zwei Jahre), Meister/Techniker, anderer Fachschulabschluss, Beamter gehobener Dienst
9 5A	Bachelor, Master, Diploma, state examination, civil servants of the highest grade	Bachelor, Master, Diplom, Magister, Staatsexamen, Beamter höherer Dienst
10 6	Doctoral degree and postdoctoral lecture qualification	Promotion

**Table 10:** Coding of CASMIN

Key	English	German
0 1a	Inadequately completed general education	Kein Abschluss
1 1b	General elementary education	Hauptschulabschluss ohne berufliche Ausbildung
2 1c	Basic vocational training above and beyond compulsory schooling	Hauptschulabschluss mit beruflicher Ausbildung
3 2b	Intermediate general education	Mittlere Reife ohne berufliche Ausbildung
4 2a	Intermediate vocational qualification, or secondary programmes in which general intermediate schooling is combined by vocational training	Mittlere Reife mit beruflicher Ausbildung
5 2c_gen	General maturity: Full maturity certificates (e.g., the Abitur, A-levels)	Hochschulreife ohne berufliche Ausbildung

(...)



**Table 10:** (continued)

Key	English	German	
6	2c_voc	Vocational maturity: Full maturity certificates including vocationally specific schooling or training	Hochschulreife mit beruflicher Ausbildung
7	3a	Lower tertiary education: Lower level tertiary degrees, generally of shorter duration and with a vocational orientation	Fachhochschulabschluss
8	3b	Higher tertiary education: The completion of a traditional, academically orientated university education	Universitätsabschluss

---

A.2 Variables affected by anonymization methods

Instrument							
Name	Additional Study Baden-Wuerttemberg						
Stage	5						
Study type	Main survey						
Study number	A72 / A73 / A74						
Dissemination	SUF						
Measures							
	data file	variable	label		On-site	RemoteNEPS	SUF-Download
<b>Countries &amp; Languages</b>							
	xTarget	t80a_g1	Country of own birth				purged
	xTarget	t84a_g1	Country of birth - mother				purged
	xTarget	t84b_g1	Country of birth - father				purged
	xTarget	t85a_g1	Country of birth of your maternal grandparents - the mother of your mother				purged
	xTarget	t85b_g1	country of birth of your maternal grandparents - the father				purged
	xTarget	t86a_g1	Country of birth of your paternal grandparents - the mother of your father				purged
	xTarget	t86b_g1	Country of birth of your paternal grandparents - the father of your father				purged
	xTarget	t80a_g2	Country of birth itself (categorized)				country aggregation
	xTarget	t84a_g2	country of birth - mother (categorized)				country aggregation
	xTarget	t84b_g2	country of birth - father (categorized)				country aggregation
	xTarget	t85a_g2	country of birth of your maternal grandparents - the mother (categorized)				country aggregation
	xTarget	t85b_g2	country of birth of your maternal grandparents - the father (categorized)				country aggregation
	xTarget	t86a_g2	country of birth of your paternal grandparents - the mother (categorized)				country aggregation
	xTarget	t86b_g2	country of birth of your paternal grandparents - the father (categorized)				country aggregation
	xTarget	t82a_g2	country of citizenship 1				country aggregation
	xTarget	t82a_g3	country of citizenship 2				country aggregation
	xTarget	t82a_g4	country of citizenship 3				country aggregation
	xTarget	t83a_g2	mother tongue 1				language aggregation
	xTarget	t83a_g3	mother tongue 2				language aggregation
	xTarget	t83a_g4	mother tongue 3				language aggregation
	xTarget	t89a_g2	mother tongue of mother 1				language aggregation
	xTarget	t89a_g3	mother tongue of mother 2				language aggregation
	xTarget	t89a_g4	mother tongue of mother 3				language aggregation
	xTarget	t90a_g2	mother tongue of father 1				language aggregation
	xTarget	t90a_g3	mother tongue of father 2				language aggregation
	xTarget	t90a_g4	mother tongue of father 3				language aggregation
	xTarget	t91_g2	other language 1				language aggregation
	xTarget	t91_g3	other language 2				language aggregation
	xTarget	t91_g4	other language 3				language aggregation
<b>String variables (Note: all string variables have been approved for reidentificational material and anonymised where necessary)</b>							
	xInstitution	e16a	Estimation further aspects of the educational reform 1				purged
	xInstitution	e16b	Estimation further aspects of the educational reform 2				purged
	xInstitution	e16c	Estimation further aspects of the educational reform 3				purged
	xInstitution	g16a	Estimation further aspects of the educational reform 1				purged
	xInstitution	g16b	Estimation further aspects of the educational reform 2				purged
	xInstitution	g16c	Estimation further aspects of the educational reform 3				purged
	xInstitution	m16a	Estimation further aspects of the educational reform 1				purged
	xInstitution	m16b	Estimation further aspects of the educational reform 2				purged
	xInstitution	m16c	Estimation further aspects of the educational reform 3				purged
	xInstitution	h14a	Estimation further aspects of the educational reform 1				purged
	xInstitution	h14b	Estimation further aspects of the educational reform 2				purged
	xInstitution	h14c	Estimation further aspects of the educational reform 3				purged
	xInstitution	h7e	Decision G8-old curricula - other, specifically				purged
	xInstitution	h28d	G9 school trial - Decision - text field Other				purged
	xInstitution	h34b	G9 school trial - Model - Other				purged
	xTarget	t40i	Reasons for G8-old/G9 switch - open entry				purged
	xTarget	t62aa	Teaching post - other - filter question				purged
	xTarget	t61a	Professional training - subject 1				purged
	xTarget	t61b	Professional training - subject 2				purged
	xTarget	t61c	Professional training - subject 3				purged
<b>Other</b>							
	xTarget	t39	Grade in which you switched to the G9 curriculum				purged
	xTarget	t97_g3	Probable place of study (RS region)				purged
	xTarget	t97_g4	Probable place of study (RS administrative district)				purged
<b>Datafiles</b>							
	xInstitution						not included

## B References

- Blossfeld, H.-P. (1985). *Bildungsexpansion und Berufschancen*. Frankfurt: Campus.
- Bundesagentur für Arbeit. (2011, March). *Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen*. Bundesagentur für Arbeit. Nuremberg.
- Bundesamt, S. (Ed.). (1992). *Klassifizierung der Berufe - Systematisches und alphabetisches Verzeichnis der Berufsbenennungen*. Stuttgart: Metzler-Poeschel.
- Ganzeboom, H. B. G. (2010). Questions and Answers about ISEI-08. Retrieved October 9, 2012, from <http://home.fsw.vu.nl/hbg.ganzeboom/isco08/qa-isei-08.htm>
- Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56.
- Ganzeboom, H. B. & Treiman, D. J. (2010). OCCUPATIONAL STATUS MEASURES FOR THE NEW INTERNATIONAL STANDARD CLASSIFICATION OF OCCUPATIONS ISCO-08; WITH A DISCUSSION OF THE NEW CLASSIFICATION.
- Müller, W., Lüttinger, P., König, W., & Karle, W. (1989). Class and Education in Industrial Nations. *International Journal of Sociology*, 19(3), 3–39.
- Office, I. L. (Ed.). (1990). *International Standard Classification of Occupations: ISCO-88*. ILO. Geneva.
- Office, I. L. (Ed.). (2012). *International Standard Classification of Occupations: ISCO-08*. Genf.
- Rohwer, G. (2011). Using Sampling Weights for Model Estimation? Retrieved October 9, 2012, from <http://www.stat.ruhr-uni-bochum.de/papers/dsw.pdf>
- Schönberger, B., Aßmann, C., & Zinn, S. (2015). *Weighting the Additonal Study Baden-Wuerttemberg of the National Educational Panel Study SUF Version 3.0.0*.
- Treiman, D. J. (1977). *Occupational prestige in comparative perspective*. New York et al: Academic Press.
- United Nations Educational, Scientific and Cultural Organization (UNESCO) (Ed.). (1997). *ISCED – International Standard Classification of Education*. United Nations Educational, Scientific and Cultural Organization (UNESCO). Paris.
- Wegener, B. (1985). Gibt es Sozialprestige? *Zeitschrift für Soziologie*, 14(3), 209–235.
- Wenzig, K. (2012). *NEPS-Daten mit DOIs referenzieren*. RatSWD Working Paper Series. Rat für Sozial- und Wirtschaftsdaten, Berlin.