Anne Weber

Data Manual

Starting Cohort 5
Context Data

DZHW : Data Manual

Dezember 2014

Table of Contents

1   Introduction

Nowadays it is not disputed that contextual factors play a significant role in educational decision making and competence development. Therefore, the National Educational Panel Study, NEPS, places special emphasis on learning environments and captures the most relevant dimensions that are expected to impact on learning and the educational career. According to a widespread distinction, educational contexts can either be formal, nonformal or informal.

Concerning the NEPS approach to assessing the formal learning environments in higher education, the primary source of information is the student's subjective perception. In addition, we also take the objectivist perspective and collect objective data mainly concerning the structural opportunities and restrictions by analyzing documents and, primarily, statistics.

Therefore, we predominantly used most recent data collected by the Federal Statistical Office and the Statistical Offices of the Länder as well as the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). In addition, we gathered relevant information published on the internet.

2   Structure of the Data File and Merging with Survey Data

We collected context data for all *413 higher education institutions* which were listed in the codebook of the Federal Statistical Office in 2010/2011.[1] Higher education institutions with different locations are only considered once with combined information for all locations.

The classification of *10 subject areas* is also similar to the classification of the Federal Statistical Office:

(1) Sports, Languages and Cultural Studies
(2) Legal Sciences
(3) Economic Sciences
(4) Social Sciences
(5) Mathematics, Natural Sciences
(6) Human Medicine / Health Sciences
(7) Veterinary Medicine / Agriculture, Forestry and Nutrition Sciences
(8) Engineering
(9) Art, Art Science
(10) Other subject areas

---

[1]   The codebook covers only officially recognized and state approved higher education institutions. Economic academies ("Wirtschaftsakademien") and universities of cooperative education ("Berufsakademien") are not included. **Therefore context data was not collected for those institutions.**

In contrast to the classification of the Federal Statistical Office we split the area "Legal, economic and social sciences" into three areas so that more detailed analyses are possible. Furthermore we combined the area "Sports" with "Languages and cultural studies" respectively "Veterinary medicine" with "Agriculture, forestry and nutrition sciences" because "Sports" and "Veterinary medicine" usually include small numbers of cases.

For ease of use, all context information has been integrated in one single datafile labeled "SC5_xInstitution_O_4-0-0.dta" (onsite) (respectively "SC5_xInstitution_R_4-0-0.dta" (remote)). This file has a multilevel structure with each subject area (variable tg04001_g7) in each higher education institution (variable ID_i). This means the context data file contains ID_i* tg04001_g7 (413*10) cases.

It is possible to merge the context data file with

- the cross-sectional information collected in the initial paper-and-pencil questionnaire (file "pTargetCATI" – main panel file CATI),
- the cross-sectional information collected in the online surveys (file "pTargetCAWI" – main panel file CAWI) and
- the longitudinal CATI module covering vocational education (file "spVocTrain").

The Stata code for merging these files is described in the following sections exemplarily for wave 1, 2 and 3. It is similar for the following waves.

2.1 Merging with the initial questionnaire (pTargetCATI)[2]

```
*open pTargetCATI-dataset
use "SC5_pTargetCATI_O_4-0-0.dta", clear

*keep only data of the first wave (data from the initial questionnaire
and CATI)
keep if wave==1

*merge with context data (matching on the key variables ID_i and
tg04001_g7)
merge m:1 ID_i tg04001_g7 using "SC5_xInstitution_O_4-0-0.dta"

*drop cases which are not in the pTargetCATI dataset (wave 1)
drop if _merge==2
```

---

[2] In the initial questionnaire, data on the higher education institution and the subject area the respondents were enrolled in at the beginning of their studies, were collected. The merging variable tg04001_g7 refers to the first subject of study.

2.2 Merging with higher education spells in the vocational training module, wave 1 (spVocTrain)[3]

```
*open spVocTrain-dataset
use "SC5_spVocTrain_O_4-0-0.dta", clear

*keep only data of the first wave
keep if wave==1

*keep only higher education spells (degree programs)
keep if (h_aktstu==1) | (h_aktstu!=1 & inlist(ts15201,6,7,8,9,10))

*fill ID_i (which is only filled, if it changed since the recruiting)
with ID_i of the Cohort Profile (which contains the recruiting infor-
mation for all cases)
merge m:1 ID_t wave using "SC5_CohortProfile_O_4-0-0.dta", keepusing
(ID_i) update

keep if inlist(_merge,3,4,5)

drop _merge

*merge with context data (matching on the key variables ID_i and
tg04001_g7)
merge m:1 ID_i tg04001_g7 using "SC5_xInstitution_O_4-0-0.dta"

*drop cases which are not in the spVocTrain-dataset (wave 1)
drop if _merge==2
```

2.3 Merging with the online survey, wave 2 (pTargetCAWI)[4]

```
*open pTargetCAWI-dataset
use "SC5_pTargetCAWI_O_4-0-0.dta", clear

*keep only data of the second wave
keep if wave==2

*drop ID_i (which is only filled, if it changed since the last tele-
phone interview) and use ID_i of the Cohort Profile (which contains the
most recent information for all cases)
drop ID_i

merge 1:1 ID_t wave using "SC5_CohortProfile_O_4-0-0.dta", keepusing
(ID_i)

keep if _merge==3

drop _merge

*merge with context data (matching on the key variables ID_i and
tg04001_g7)
merge m:1 ID_i tg04001_g7 using "SC5_xInstitution_O_4-0-0.dta"

*drop cases which are not in the pTargetCAWI-dataset (wave 2)
drop if _merge==2
```

---

[3] The merging variable tg04001_g7 refers to the first subject of study.
[4] The merging variable tg04001_g7 refers to the "reference subject of study" which is identified during the online survey.

2.4 Merging with higher education spells in the vocational training module, wave 3 (spVocTrain)[5]

```
*open spVocTrain-dataset
use "SC5_spVocTrain_O_4-0-0.dta", clear

*keep only data of the third wave
keep if wave==3


*keep only higher education spells (degree programs)
keep if (h_aktstu==1) | (h_aktstu!=1 & inlist(ts15201,6,7,8,9,10))

*fill primary keys with information of the preceding episode if it is
the same
bysort ID_t: replace tg04001_g7=tg04001_g7[_n-1] if tg04001_g7==. &
tg24159==1

bysort ID_t: replace ID_i=ID_i[_n-1] if ID_i==. & tg24121==1

*merge with context data (matching on the key variables ID_i and
tg04001_g7)
merge m:1 ID_i tg04001_g7 using "SC5_xInstitution_O_4-0-0.dta"

*drop cases which are not in the spVocTrain dataset (wave 3)
drop if _merge==2
```

3 Reference Level and Variables

In the context data file 118 variables are included. Besides the primary keys ID_i and tg04001_g7, there are 116 variables with context information.

Context data is offered on the following three levels:

(1) Local, regional context

The first level provides information about the local, regional context of a higher education institution, e.g. the settlement structure and the economic and social structure. For higher education institutions with different locations the information on the local/regional context level refers to the location with the highest number of students.

(2) Higher education institution

The second level includes information about a higher education institution itself, for instance its type and its institutional control or the number of students.

(3) Subject area

The third level contains information about each subject area in a higher education institution, e.g. number of students, number of academic staff or the financial resources.

Table 1 presents an overview of the indicators we generated on each level and the information sources we used.

---

5   The merging variable tg04001_g7 refers to the first subject of study.

Notes:

- *Systemmissings mean that the information was not available in the used sources.*
- *Negative values in the variables relating to financial resources (tg93301, tg93302, tg93501) stand for an overplus.*
- *In the variables relating to human resources in the different subject areas (tg93401 to tg93430), data was deleted by an anonymization software of the Federal Statistical Office if there was only a small number of cases. Therefore, inconsistencies between the sum of the seperate data for men/women and the total number may occur.*

Table 1: Variables, sources and reference dates[6]

ID_i	**Institution-ID**	**Federal Statistical Office**	**12/31/2010**
tg04001_g7	**Subject area of studies**	**Federal Statistical Office (Students at Higher Education Institutions, Main Reports)**	**12/31/2010**

(1) Local, regional context

Settlement structure

BIK-region

tg91102	**BIK-region type**	**Federal Statistical Office (Register of Municipalities/GV100NAD)**	**12/31/2009**
tg91103	BIK-structure type	Federal Statistical Office (Register of Municipalities/GV100NAD)	12/31/2009
tg91104	BIK-7 class of municipality size	Federal Statistical Office (Register of Municipalities/GV100NAD)	12/31/2009
tg91105	BIK-10 class of municipality size	Federal Statistical Office (Register of Municipalities/GV100NAD)	12/31/2009

BBSR-region

tg91106	BBSR community type	BBSR (Continuous Spatial Observation (www.bbsr.bund.de))	12/31/2008
tg91107	BBSR district type	BBSR (INKAR-CD 2011)	12/31/2009
tg91108	BBSR composite district type	BBSR (INKAR-CD 2011)	12/31/2009
tg91109	BBSR basic region type	BBSR (INKAR-CD 2011)	12/31/2009
tg91110	BBSR differentiated basic region type	BBSR (INKAR-CD 2011)	12/31/2009

---

[6] **The onsite datafile includes all variables. Variables which are (anonymized) also in the remote datafile are marked in bold.**

Population density

tg91111	Population density 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91112	Population-employment-density 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91113	Population density 2009 in the regional planning area	BBSR (INKAR-CD 2011)	12/31/2009
tg91114	Population-employment-density 2009 in the regional planning area	BBSR (INKAR-CD 2011)	12/31/2009

Labor market

tg91201	**Unemployment rate 2009 in the county**	**BBSR (INKAR-CD 2011)**	**12/31/2009**
tg91202	Proportion long-term unemployed 2008 in the county	BBSR (INKAR-CD 2011)	12/31/2008
tg91203	Proportion low-skilled employees 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91204	Proportion high-skilled employees 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009

Social structure

tg91301	Proportion Population beyond 15 y. with university degree 2009 in NUTS2-region	BBSR (INKAR-CD 2011)	12/31/2009

Economic situation

tg91401	Gross domestic product per inhabitant 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91402	Monthly household income 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91403	Community debts per inhabitant 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009

Political-administrative structure

tg91502	**West Germany/East Germany/Berlin**	**Federal Statistical Office**	**12/31/2009**

Economic structure

Employees in economic sectors

tg91601	Proportion employees in FuE 2007 in the county	BBSR (INKAR-CD 2011)	12/31/2007
tg91602	Proportion employees in secondary sector 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009

tg91603	Proportion employees in the tertiary sector 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91604	Service rate 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91605	Proportion employees enterprise related services 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91606	Proportion employees in creative sectors 2009 in the county	BBSR (INKAR-CD 2011)	12/31/2009
tg91607	Average number employees in agriculture/wood/fishing 2009 in the county	Federal Statistical Office (Regional Database)	12/31/2009
tg91608	Average number employees in manifacturing industry without building industy 2009 in the county	Federal Statistical Office (Regional Database)	12/31/2009
tg91609	Average number employees in building industry 2009 in the county	Federal Statistical Office (Regional Database)	12/31/2009
tg91610	Average number employees in trade/customer/traffic 2009 in the county	Federal Statistical Office (Regional Database)	12/31/2009
tg91611	Average number employees in financing/letting/business services 2009 in the county	Federal Statistical Office (Regional Database)	12/31/2009
tg91612	Average number employees in public/private services 2009 in the county	Federal Statistical Office (Regional Database)	12/31/2009
Number of current businesses			
tg91613	Number of current businesses with 0-9 employees 2009 in the county	Federal Statistical Office (Regional Database [Business Register System 95 (URS95)])	12/31/2009
tg91614	Number of current businesses with 10-49 employees 2009 in the county	Federal Statistical Office (Regional Database [Business Register System 95 (URS95)])	12/31/2009
tg91615	Number of current businesses with 50-249 employees 2009 in the county	Federal Statistical Office (Regional Database [Business Register System 95 (URS95)])	12/31/2009
tg91616	Number of current businesses with 250+ employees 2009 in the county	Federal Statistical Office (Regional Database [Business Register System 95 (URS95)])	12/31/2009

(2) Higher education institution

Orientation as elite university

tg92101	Nomination future concept 2006 or 2007	www.dfg.de	12/31/2007
tg92102	Winner future concept 2006 or 2007	www.dfg.de	12/31/2007
tg92103	Nomination Cluster of excellence 2006 or 2007	www.dfg.de	12/31/2007
tg92104	Winner Cluster of excellence 2006 or 2007	www.dfg.de	12/31/2007
tg92105	Nomination graduate school 2006 or 2007	www.dfg.de	12/31/2007
tg92106	Winner graduate school 2006 or 2007	www.dfg.de	12/31/2007

Funding body

tg92301	**Funding body**	**www.hochschulkompass.de; Federal Statistical Office (Students at Higher Education Institutions, Main Reports)**	**12/31/2009**

Type of higher education institution

tg92401	**Type of higher education institution**	**Federal Statistical Office**	**12/31/2010**

Quality of teaching

tg92501	Finalist Excellent teaching 2009	www.exzellente-lehre.de	12/31/2009
tg92502	Winner Excellent teaching 2009	www.exzellente-lehre.de	12/31/2009

Size

Students total

tg92601	**Students 2010 total**	**Federal Statistical Office (Students at Higher Education Institutions, Main Reports)**	**12/31/2010**

Beginning students total

tg92602	Beginning students 2010 (1st semester) total	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

Social composition

Students

tg92701	Students 2010: Germans	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92702	Students 2010: Foreign students with a German school leaving certificate qualifying for higher education („Bildungsinländer")	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92703	Students 2010: Foreign students	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92704	Students 2010: male	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92705	Students 2010: female	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

Beginning students

tg92706	Beginning students 2010 (1st semester): Germans	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92707	Beginning students 2010 (1st semester): "Bildungsinländer"	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92708	Beginning students 2010 (1st semester): Foreign students	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92709	Beginning students 2010 (1st semester): male	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92710	Beginning students 2010 (1st semester): female	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

Composition of subject areas

tg92801	Students 2010: Subject area sports/language/cultural sciences	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92802	Students 2010: Subject area Law	Federal Statistical Office (Students at Higher Educa-	12/31/2010

		tion Institutions, Main Reports)	
tg92803	Students 2010: Subject area Business	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92804	Students 2010: Subject area Social sciences	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92805	Students 2010: Subject area mathematics/natural sciences	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92806	Students 2010: Subject area medicine/health sciences	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92807	Students 2010: Subject area veterinary/agricultural/forestry/nutrition	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92808	Students 2010: Subject area engineering	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92809	Students 2010: Subject area art/art sciences	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg92810	Students 2010: Other subjects	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

(3) Subject area

Size

Number of students

tg93101	Students 2010	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

Number of beginning students

tg93102	Beginning students 2010 (1st semester) 2010	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

Social composition

Students

tg93201	Students 2010: Germans	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg93202	Students 2010: "Bildungsinländer"	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg93203	Students 2010: Foreign students	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg93204	Students 2010: male	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg93205	Students 2010: female	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

Beginning students

tg93206	Beginning students 2010 (1st semester): Germans	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg93207	Beginning students 2010 (1st semester): "Bildungsinländer"	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg93208	Beginning students 2010 (1st semester): Foreign students	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

tg93209	Beginning students 2010 (1st semester): male	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010
tg93210	Beginning students 2010 (1st semester): female	Federal Statistical Office (Students at Higher Education Institutions, Main Reports)	12/31/2010

Financial resources

tg93301	Regular basic resources 2010	Federal Statistical Office (Financial Statistics for Higher Education Institutions)	12/31/2010
tg93302	External funds for teaching and research in general 2010	Federal Statistical Office (Financial Statistics for Higher Education Institutions)	12/31/2010

Human resources

Number of professors

tg93401	Number of professors	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93402	Number of professors: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93403	Number of professors: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93404	Full-time equivalent professors	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93405	Full-time equivalent professors: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93406	Full-time equivalent professors: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010

Number of lecturers, assistants, scientific/artistic employees

tg93407	Number of lecturers, assistants, academic and artistic staff	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93408	Number of lecturers, assistants, academic and artistic staff: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010

tg93409	Number of lecturers, assistants, academic and artistic staff: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93410	Full-time equivalent lecturers, assistants, academic and artistic staff	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93411	full-time equivalent lecturers, assistants, academic and artistic staff: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93412	Full-time equivalent lecturers, assistants, academic and artistic staff: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010

Number of lecturers, assistants, scientific/artistic employees without external funds

tg93413	Number of lecturers, assistants, academic and artistic staff without third-party funded staff	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93414	Number of lecturers, assistants, academic and artistic staff without third-party funded staff: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93415	Number of lecturers, assistants, academic and artistic staff without third-party funded staff: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93416	Full-time equivalent lecturers, assistants, academic and artistic staff without third-party funded staff	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93417	Full-time equivalent lecturers, assistants, academic and artistic staff without third-party funded staff: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93418	Full-time equivalent lecturers, assistants, academic and artistic staff without third-party funded staff: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010

Number of teaching staff for special tasks

tg93419	Number of teaching staff for special tasks	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93420	Number of teaching staff for special tasks: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93421	Number of teaching staff for special tasks: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010

tg93422	Full-time equivalent teaching staff for special tasks	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93423	Full-time equivalent teaching staff for special tasks: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93424	Full-time equivalent teaching staff for special tasks: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010

Number of part-time lecturers

tg93425	Number of part-time lecturers	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93426	Number of part-time lecturers: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93427	Number of part-time lecturers: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93428	Full-time equivalent part-time lecturers	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93429	Full-time equivalent part-time lecturers: male	Federal Statistical Office (special analysis upon inquiry)	12/31/2010
tg93430	Full-time equivalent part-time lecturers: female	Federal Statistical Office (special analysis upon inquiry)	12/31/2010

Research performance

tg93501	**External funds per professor**	**Federal Statistical Office (calculation based on tg93302 und tg93401)**	**12/31/2010**

Student-teacher ratio

tg93601	**Students per professor**	**Federal Statistical Office (calculation based on tg93101 und tg93401)**	**12/31/2010**
tg93602	**Students per lecturers, assistants, academic and artistic staff without third-party funded staff (full-time equivalent)**	**Federal Statistical Office (calculation based on tg93101 und tg93416)**	**12/31/2010**