Weighting in PIAAC-L 2015

Burkhardt, Luise; Bartsch, Simone

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

gesis Leibniz-Institut für Sozialwissenschaften

GESIS *Papers* 2017|30

Weighting in PIAAC-L 2015

Luise Burkhardt & Simone Bartsch

German Socio-Economic Panel Study (SOEP),
DIW Berlin, Germany

GESIS-Papers 2017|30

Weighting in PIAAC-L 2015

Luise Burkhardt & Simone Bartsch

German Socio-Economic Panel Study (SOEP),
DIW Berlin, Germany

GESIS – Leibniz-Institut für Sozialwissenschaften 2017

GESIS-Papers

GESIS – Leibniz-Institut für Sozialwissenschaften
Survey Design and Methodology
Postfach 12 21 55
68072 Mannheim
Telefon: (0621) 1246 - 518
Telefax: (0621) 1246 - 100
E-Mail: anouk.zabal@gesis.org

Correspondence: Luise Burkhardt, German Socio-Economic Panel Study (SOEP), DIW Berlin, Mohrenstraße 58, Berlin, Germany, Phone: +49 30 89789-438, lburkhardt@diw.de

Acknowledgments: The authors would like to thank the members of the project team, in particular Silke Martin and Anouk Zabal, for their continuous support as well as for their help as reviewers. Further special thanks to Tobias Silbermann for his indispensable work.

PIAAC-L was funded by the Federal Ministry of Education and Research, Berlin (Grant number 01 JP 1301 A, B, C)

ISSN: 2364-3781 (Online)
Herausgeber,
Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Contents

1 Introduction

In 2011/2012, key adult competencies were assessed in 24 countries (including Germany) as a part of the OECD Programme for the International Assessment of Adult Competencies, PIAAC (Zabal et al., 2014). In order to enrich the analytical power of the PIAAC data, the German PIAAC-Longitudinal Project (PIAAC-L)[1] follows up the original German PIAAC 2012 respondents that could be re-contacted, as well as members of their households, ages 18 and over, with three additional waves of data collection (in 2014, 2015, and 2016). This study is a cooperative project of GESIS – Leibniz Institute for the Social Sciences, the National Educational Panel Survey (NEPS) at the Leibniz Institute for Educational Trajectories (LIfBi), and the Socio-Economic Panel (SOEP) at DIW Berlin and combines research questions and measurement instruments from all three institutes (Zabal, Martin, & Rammstedt, 2016; Rammstedt, Martin, Zabal, Carstensen, & Schupp, 2017).

The present paper describes the weighting process for the second of the three PIAAC-L waves. The weighting process for the first PIAAC-L wave is outlined in Bartsch, Poschmann, and Burkhardt (2017). As weighting in PIAAC-L follows a consistent concept across waves, parts of the present paper were taken literally from Bartsch et al. (2017). Where applicable, updates or adaptations were undertaken, based on wave 2 weighting processes and data from dataset ZA5989_Weights_15. This dataset is one of seven sub-datasets that were released for 2015 as part of the PIAAC-L database that encompasses data from 2014 and 2015.[2, 3]

We start with a short introduction to some features of PIAAC-L that are important to understand the weighting procedure in PIAAC-L in general. Subsequently, we illustrate the weighting procedure specifically applied in PIAAC-L 2015 by describing the two weighting steps nonresponse adjustment and post-stratification or calibration.

[1] Commissioned by the Federal Ministry of Education and Research, Berlin, Grant number 01 JP 1301 A, B, C.

[2] Last update 22.02.2017: GESIS – Leibniz Institute for the Social Sciences, German Socio-Economic Panel (SOEP) at DIW Berlin & LIfBi – Leibniz Institute for Educational Trajectories (2017): PIAAC-Longitudinal (PIAAC-L), Germany. GESIS Data Archive, Cologne. ZA5989 Data file Version 2.1.0, doi: 10.4232/1.12734.

[3] The PIAAC-L 2016 datasets will be released in December 2017 (GESIS – Leibniz Institute for the Social Sciences, German Socio-Economic Panel (SOEP) at DIW Berlin & LIfBi – Leibniz Institute for Educational Trajectories (2017): PIAAC-Longitudinal (PIAAC-L), Germany. GESIS Data Archive, Cologne. ZA5989 Data file Version 3.0.0, doi: 10.4232/1.12925.). This release will also contain updates of the PIAAC-L 2014 and 2015 datasets; there is no update for the dataset ZA5989_Weights_15.

2 Weighting in Panel Studies: The PIAAC-L Weighting Concept

One major challenge of the PIAAC-L project was to harmonize different approaches in PIAAC and the SOEP. Compared to more "regular" panel studies like e.g. the SOEP (Wagner, Frick, & Schupp, 2007; Kroh, Siegers, & Kühne, 2015) that pursue a panel survey design from scratch, the recruitment of PIAAC participants–here referred to as PIAAC 2012 anchor persons–into a panel survey (PIAAC-L) has followed a somewhat different path (see Zabal et al., 2016; Bartsch et al., 2017). Hence one consequence is that in PIAAC-L weighting was only performed for PIAAC 2012 anchor persons and not for other household members who participated in PIAAC-L only. There is no data available to calculate the sampling probability of these household members, and thus weighting would have had to follow a different approach for this group, which was not further pursued.[4] As only the PIAAC 2012 anchor person is followed-up in the three waves of PIAAC-L, this is also plausible from an analytical perspective.

For PIAAC 2012, a cross-sectional weight and replicate weights were calculated.[5] These weights include both a nonresponse analysis and a post-stratification. In a panel study like the SOEP, nonresponse weights are calculated separately to account for the panel attrition and also to enable separate analyses of single waves. The nonresponse weight flows into the calculation of the cross-sectional weight. Since PIAAC-L has transformed into a panel study, the panel approach was applied here as well and both, nonresponse weight (bleib) and cross-sectional weight (hrf), were calculated and delivered.

As described in Bartsch et al. (2017), for weighting in PIAAC-L 2014 the nonresponse analysis was split into four separate models, taking into account four reasons for attrition on the way from PIAAC 2012 to PIAAC-L (literacy-related nonresponse; nonresponse due to assessment break-off or non-consent; nonresponse due to noncontact and general nonresponse). In the case of PIAAC-L 2015, we are dealing with a rather "normal" follow-up wave within a panel study. Thus, the respondents are already informed about the fact that they will be re-contaced for another interview in the following year, the address is validated within the field process and thus, we only distinguish the two classical steps for nonresponse analysis: modelling of noncontact and modelling of nonresponse. Please note that in PIAAC-L 2015 not all household members were eligible. Interviews were only conducted with PIAAC 2012 anchor persons and their spouses or partners living in the same household. Weighting was again only performed for PIAAC 2012 anchor persons and not for their partners or spouses.

In brief: variance estimation

Since PIAAC-L is a follow-up study to PIAAC 2012, addressing German PIAAC respondents who had consented to being re-contacted, the starting point with regard to sampling is the original sample selection in PIAAC. Thus, for purposes of variance estimation, users should use variables on sampling and stratification provided in the PIAAC 2012 scientific use file, such as the variables VARSTRAT, VARUNIT, ID_PSU, STRAT_PSU, Federal_state, or GKPOL. Replicate weights for variance estimation, as provided for the PIAAC scientific use file, are not computed for PIAAC-L.

The next section describes the calculation of the nonresponse weight (bleib_15) for PIAAC-L 2015.

[4] See Saßenroth, Kroh, and Wagner (2013) for a weighting approach in which no sampling probability can be calculated.

[5] Readers not familiar with the PIAAC 2012 weighting procedure are referred to Zabal et al. (2014), pp. 80 and Mohadjer, Krenzke, and Van de Kerckhove (2013) for further information.

3 Modelling Nonresponse and Nonresponse Weights

The nonresponse analysis in PIAAC-L 2015 was split into two parts: modelling of noncontact and modelling of nonresponse. For each of the two steps a logistic regression model is estimated, where the dependent variable is a 0/1-variable (nonresponse/response) and the independent variables were selected as explained below. The predicted probabilities (of contact and response respectively) are derived from each of these two models and the inverse product of these probabilities yields the nonresponse weights.

3.1 Dependent Variables and Response Rate

As mentioned earlier, in PIAAC-L 2015 only anchor persons who participated in the 2014 data collection and their partners (living in the same household) were eligible for participation. As weighting was only performed for anchor persons and not for their partners or spouses, the gross sample for weighting in 2015 (i.e. field phase in 2015) is comprised of n=3,758 eligible anchor persons. This gross sample was further divided into three groups: nonresponse due to noncontact (M1) and nonresponse due to refusal or due to other reasons (e.g. long-term illness, linguistic problems) (M2) and of course respondents (see Table 1).

Table 1 describes the results of the fieldwork for PIAAC-L 2015. The adjusted response rate, i.e. the number of valid interviews with eligible persons (n=3,263) divided by the number of eligible persons in the gross sample for fieldwork excluding neutral (ineligible) dropouts (n=3,746) was 87.1%. The unadjusted response rate, referring to the gross sample including neutral (ineligible) drop-outs, amounts to 86.8% (Table 1 and Steinacker & Wolfert, 2017, p. 38).

A nonresponse rate of 12.9% of the gross sample including nonresponse due to noncontact may lead to a nonresponse bias, which should be corrected for (Groves & Peytcheva, 2008). With respect to the net sample of PIAAC-L 2014, which included 3,758[6] cases, the analysis was undertaken in two steps. Such a differentiation between nonresponse due to noncontact and nonresponse due to refusal or other reasons is appropriate and is performed in established panel studies like the SOEP (see Kroh, 2014).

[6] Please note that numbers in Table 2 differ because neutral (ineligible) dropouts, i.e. people who moved abroad or deceased (N=12) were not included in the analyses, following the standard weighting approach of the SOEP (Kroh, Käppner, & Kühne, 2014).

Table 1 Result of Fieldwork for PIAAC-L 2015

Final results	abs.	%
Interview		
Interview valid	3,263	86.8
Ineligible		
Anchor person moved abroad	6	0.2
Anchor person deceased	6	0.2
Noncontact (M1)		
Declined participation before start of fieldwork	7	0.2
Anchor person moved to unknown address	28	0.7
Anchor person moved to known address	2	0.1
Address no longer exists	3	0.1
Anchor person unknown at given address	1	0.0
No one home	37	1.0
Nonresponse (M2)		
Interview impossible during fieldwork	63	1.7
Anchor person unable to respond due to long-term illness or other reason	8	0.2
Linguistic problems, inadequate German skills	3	0.1
Unwilling to participate in interview	237	6.3
Correct address, but anchor not met in person	9	0.2
Other reasons, unusual circumstances	83	2.2
Contact established without final result	2	0.1
Gross sample	3,758	100.0[7]

3.2 Independent Variables

The aim of modelling nonresponse is to produce response propensities that can serve as a basis for weighting factors and compensate for attrition. To model nonresponse, variables are needed that are available for both, respondents and nonrespondents. In the case of a panel or follow-up survey, like PIAAC-L, characteristics from previous waves can be used to explain potential selectivity in the following waves. To calculate weights for PIAAC-L 2015, a number of variables from PIAAC-L 2014 were included; in addition, information on the native language of the respondents was taken from the PIAAC 2012 survey. Furthermore, information from the PIAAC-L 2015 fieldwork was used (e.g., information on interviewer changes between waves and mode of initial contact). For each of the steps, the selection of explanatory variables was based on established assumptions and theories in the field of survey methodology concerning their power to explain nonresponse (also see Kroh, Käppner, & Kühne, 2014) and was also aligned with the selection of explanatory variables that were used to model nonresponse in PIAAC-L 2014. For details on the final set of variables see Table 2.

[7] Please note: For the sake of clarity, values in this paper are rounded, thus the percentages presented in these tables may not add up exactly to 100.0.

Table 2 Summary of Independent Variables

	Model 1 Noncontact	Model 2 Nonresponse
Interview Characteristics		
Interviewer change (from wave 1 to wave 2)	X	X
Initial contact via phone		X
Partial unit nonresponse (nonresp. of eligible hh-member last wave)		X
Address change	X	X
Region and Household Characteristics		
Federal state	X	X
Size of municipality (grouped)		X
Housing / building type	X	X
Home owner	X	
Subtenant	X	
Household size		X
Income and Employment Characteristics		
Income (quartiles)	X	X
Labor force participation		X
Occupational position		X
Job change	X	
Family and Partnership Characteristics		
Partner (currently in relationship)		X
Married	X	
Divorced		X
Family member(s) in household		X
Children in household	X	X
Education Characteristics		
Education (school degree, grouped)		X
Education parents (school degree, grouped)	X	X
Other Characteristics		
Age (grouped)[8]	X	X
German citizenship	X	X
Native language German	X	X
Disability		X
Cognitive Skills Characteristics		
Literacy (quartiles)	X	
Numeracy (quartiles)	X	
Problem solving (quartiles)		X
n	3,746	3,668

[8] This refers to the anchor person's age at the time of the 2014 PIAAC-L interview (range: 18-68). In the post-stratification process, however, the actual age as reported by the respondent in the PIAAC-L 2015 interview is used for calibration.

Modelling of nonresponse aims at consistent estimation of response propensities. There is no focus on an interpretation of the effects (Spieß, 2010). All independent variables were checked for association with the outcome variables—noncontact and nonresponse respectively—and those with a statistically significant association at the 10%-level were included in a full model. This full model was subsequently reduced to those variables significant at the 5%-level. Due to the fact that the selection of the independent variables is data-driven and not theory-driven, apparently important explanatory variables such as gender, education, labor force participation and so on are not included in either both or at least one reduced model as they showed no significance at the 5%-level.

In PIAAC 2012, for each assessed competency domain, literacy, numeracy and problem solving in technology-rich environments, 10 plausible values were computed per respondent (Yamamoto, Khorramdel, & von Davier, 2013). For weighting in PIAAC-L 2015, the re-scaled proficiency measures from PIAAC 2012 were used (for details on re-scaling see the accompanying data documentation *Notes to the User)*[9]. The re-scaled measures are provided in the updated dataset ZA5989_Persons_14[10] and were used in the nonresponse adjustment models by first calculating the mean across all ten plausible values for each domain. In a second step, quartiles were calculated. For the competency domain problem solving, an extra missing category was calculated for cases in which no plausible values were available (n=527).[11]

3.3 Model 1 – Noncontact

Model 1 refers to step one of the analysis: predicting the probability for a successful re-contact in relation to the probability of an unsuccessful attempt to re-contact the respondents from the prior PIAAC-L 2014 wave. Hence, it includes all variables that were considered for the attrition analysis and showed significant association to the contact indicator at a 10%-level ($p<0.10$). In this step an attrition of 78 cases was recorded. As mentioned earlier, first a full model with all of these variables was calculated and then stepwise routines were run in order to calculate a reduced model with only significant factors at the 5%-level ($p<0.05$). Table 3 and Figure 1 show the outcomes of the full and reduced model. Figure 1 shows the coefficients and the 95%-confidence intervals for illustration. It can be seen that only few variables remain in the reduced model, which has a rather low statistical power.

[9] Accessible under https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5989&tsearch=piaac&tsearch2=&tfield=all&tfield2=&tdb=e&ttab=0&tnotabs=&tnf=1&taf=&tll=10

[10] Last update 22.02.2017: GESIS – Leibniz Institute for the Social Sciences, German Socio-Economic Panel (SOEP) at DIW Berlin & LIfBi – Leibniz Institute for Educational Trajectories (2017): PIAAC-Longitudinal (PIAAC-L), Germany. GESIS Data Archive, Cologne. ZA5989 Data file Version 2.1.0, doi: 10.4232/1.12734.

[11] No plausible values were available for respondents who did a paper-based assessment in PIAAC 2012 because the paper instruments did not include the competency domain problem solving.

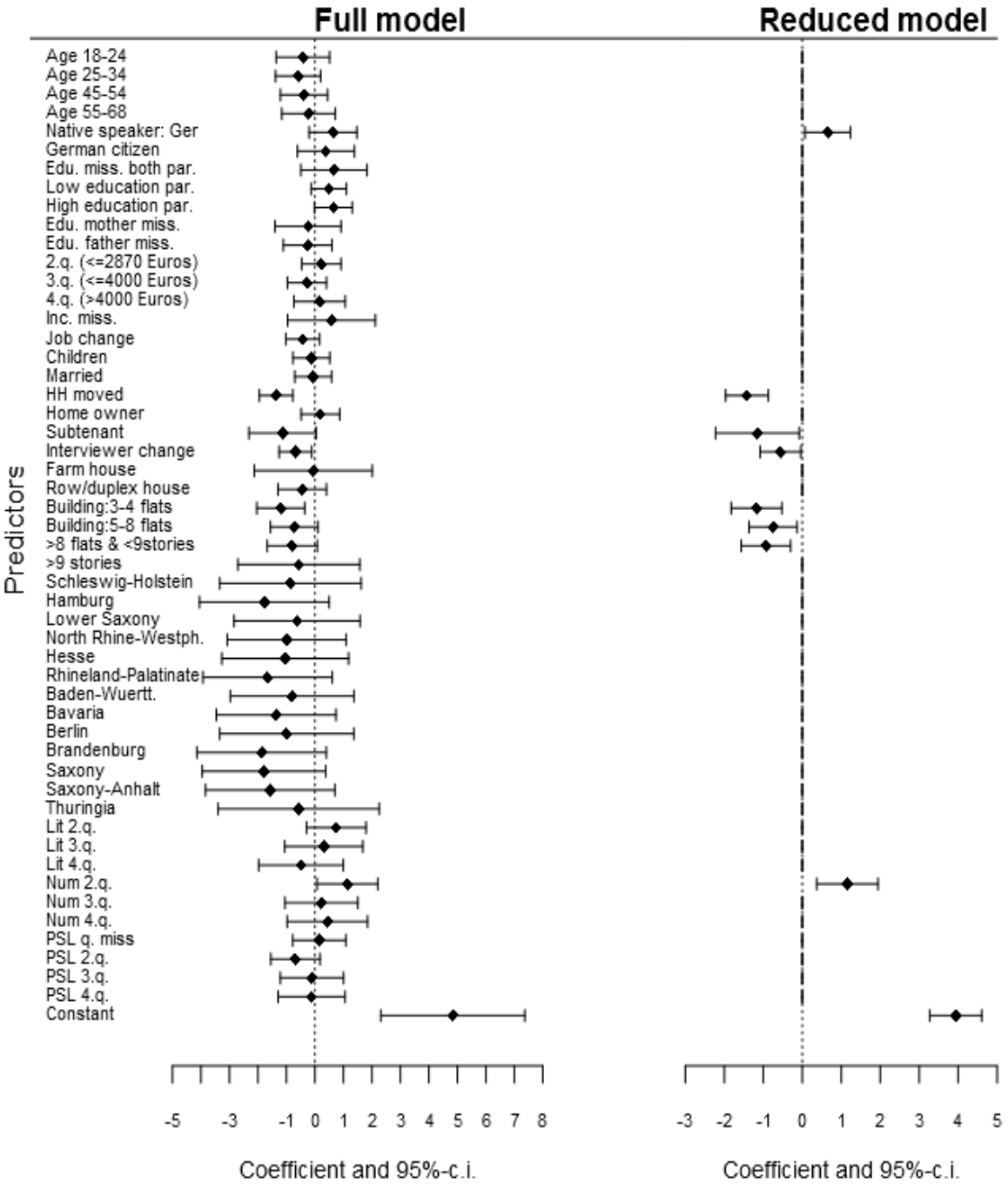

Figure 1 Coefficients and Confidence Intervals for Model 1 (Noncontact)[12]

[12] For a list of abbreviations used in Figure 1, see Table A1 in the appendix.

Table 3 Fit Values for Estimated Models for Step 1

	Full model	Reduced model
Observations	3,746	3,746
Pseudo-R^2	0.145	0.094
R^2 (McFadden)	0.008	0.070

3.4 Model 2 – Nonresponse

Model 2 refers to step two of the analysis: The final nonresponse to the request of the interviewer to take part in PIAAC-L 2015 due to refusal or due to other reasons such as long-term illness. This is, in numerical terms, the biggest step in the attrition process, with 405 individuals refusing to participate again. Still, the resulting statistical power of the reduced model is again rather low (see Table 4). Even though more variables from the PIAAC-L 2014 wave itself remain in the model, only a small portion of the variables shows statistically significant explanatory power at the 5%-level, and those left in the reduced model still show confidence bounds close to zero (see Figure 2). As mentioned earlier, the content-related interpretation of the model is not in the focus when modelling nonresponse.

Table 4 Fit Values for Estimated Models for Step 2

	Full Model	Reduced Model
Observations	3,668	3,668
Pseudo-R^2	0.076	0.059
R^2 (McFadden)	0.025	0.044

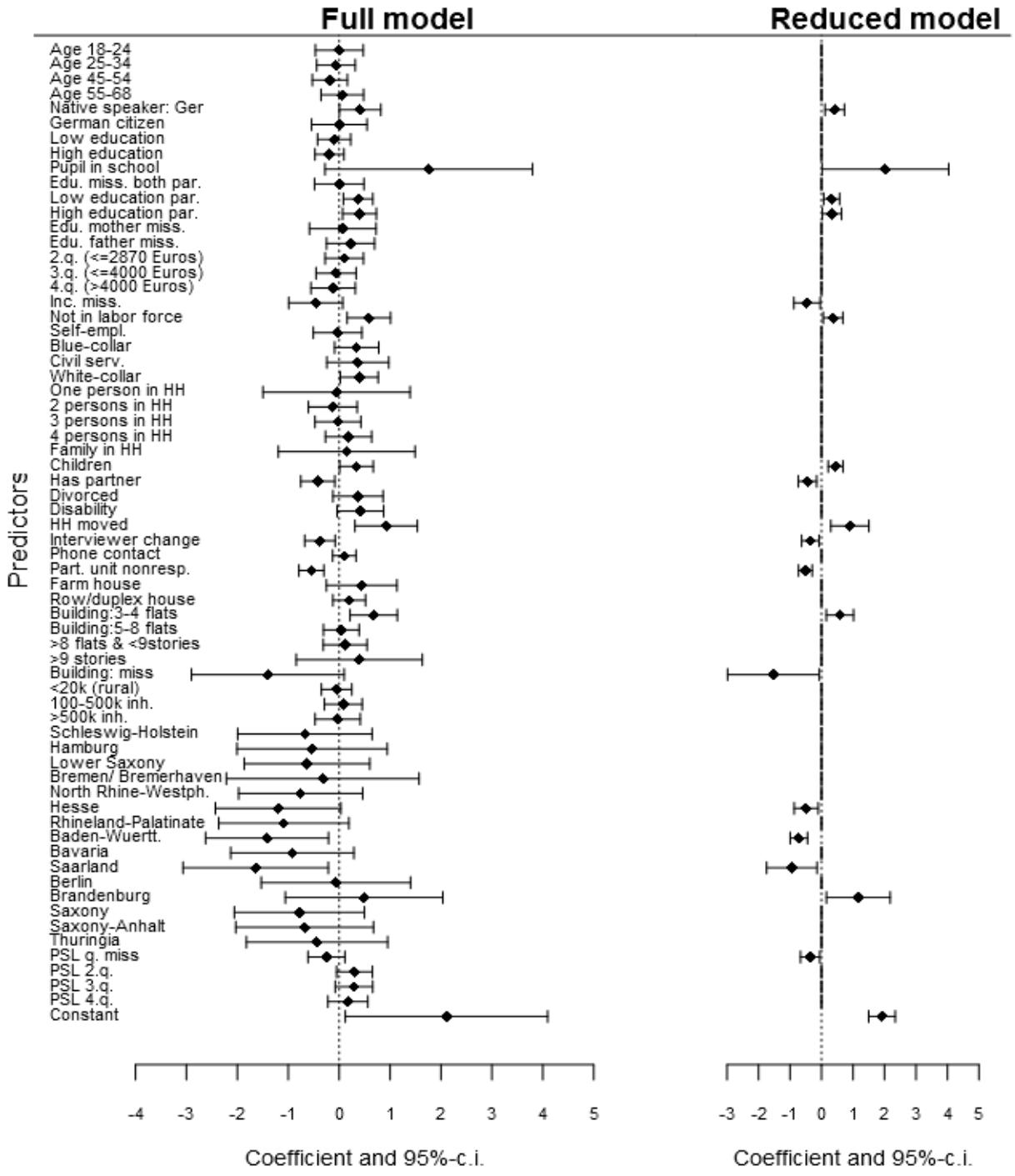

Figure 2 Coefficents and Confidence Intervals for Model 2 (Nonresponse)[13]

[13] For a list of abbreviations used in Figure 2, see Table A1 in the appendix.

3.5 Final Nonresponse Weights

To calculate the nonresponse weight, the staying probability for each of the two models was calculated. The staying probability (P(WB=1)) is the product of the probability to not participate in the interview due to noncontact and the probability to not participate in the interview due to nonresponse (nonresponse due to refusals or other reasons): P(WB=1) = P(M1=1)*P(M2=1). The inverse of the staying probability yields the raw nonresponse weight.

In PIAAC 2012 as well as in the SOEP, weights were trimmed when exceeding certain thresholds. In PIAAC-L 2014 the nonresponse-weights were also trimmed when exceeding twice the median (see Bartsch et al., 2017). The selection of the cut-off point is basically driven by the decision not to trim more than 1% of the derived weights and thus maintain efficacy.

Table 5 Estimated Nonresponse Weights (bleib_15)

	Min	10%	50%	75%	Max	Mean	SD
Final nonresponse weights	1.006	1.067	1.124	1.180	1.875	1.148	0.096

Since the nonresponse weights in PIAAC-L 2015 showed only low dispersion, no trimming procedure had to be applied (see Table 5). The distribution shows an elongated right tail (see Figure 3).

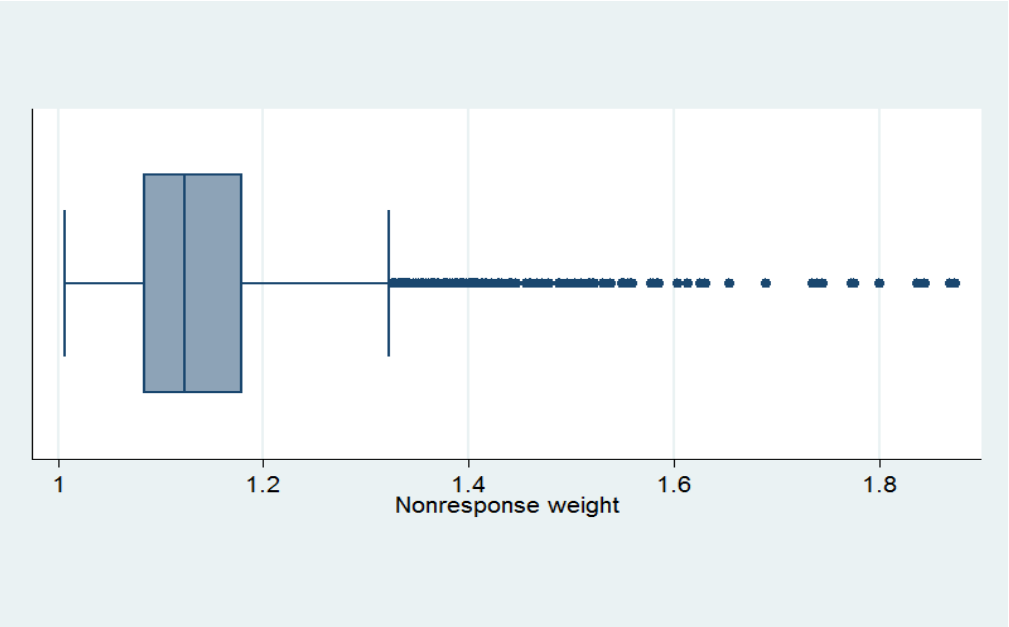

Figure 3 Distribution of Estimated Nonresponse Weights (bleib_15)

4 Calibration and Delivered Weights

The process of calculating weights described so far reflects the nonresponse adjustment. Calibration, however, aims at bringing the sample in closer alignment with the underlying population, at least with regard to the distribution of some central variables. This is generally done by using data from official statistical sources. In the case of Germany, the Microcensus is the source for the reference data.

Since the PIAAC-L sample had no refreshment of sample members, the reference population is described as *non-institutionalized adults born between November 1946 and November 1995 that did not move to Germany in 2012 or later* and is the same as the reference population for PIAAC 2012. For this reference group, a separate count of the most up-to-date data from the 2015 Microcensus (Statistisches Bundesamt, 2016) serves as the basis for calibration. This means that the population to which the weights refer consists of 51.40 million persons (Statistisches Bundesamt, 2016).[14]

Considering the reduced sample size as a product of various factors, it was not possible to replicate one-to-one the PIAAC post-stratification process. Instead, the PIAAC-L post-stratification process implemented a mixed approach between raking and post-stratification (see Bartsch et al., 2017).

In PIAAC-L 2015 the combined table for raking used the variables gender (2 categories), age (5), and education (3)[15] as raking references at the individual level and the variables region (3), size of household (5), and size of municipality (7) as additional variables at the household level.[16] The additional variables were used according to the weighting approach of the German Socio-Economic Panel (Kroh et al., 2014), keeping in mind that the number of variables adjusted for should still be kept to a minimum.

The basis for the raking procedure was the product of the nonresponse adjustment factor (bleib_15) and the cross-sectional weight from PIAAC-L 2014 (hrf_14). The descriptive statistics of the resulting weights are given in Table 6.

Table 6 Descriptive Statistics of the Final PIAAC-L 2015 Weights

	Min	10%	50%	75%	Max	Mean	SD
Final weights	2836.07	7374.35	13257.11	18796.22	95043.97	15753.60	9572.53

[14] Please note: Due to its design it could not be ensured that the sample of the Microcensus precisely resembles the structure of the PIAAC-L sample, which, by definition, only includes persons in non-institutionalized households–both in 2011 and 2015. The reference sample might include a presumably small number of people that lived in institutions before 2015 as the Microcensus is a cross-sectional survey.

[15] Since the subsample of pupils in school naturally decreases over time, the categories "Still in school" and "Low educational level" were combined for calibration in PIAAC-L 2015. Thus, the number of categories for education was reduced from four to three compared to PIAAC 2012 and PIAAC-L 2014.

[16] For all variables, up-to-date information from 2015 was used.

4.1 Usage of Weights

The dataset ZA5989_Weights_15 includes the two weighting factors bleib_15 and hrf_15. The weighting factor hrf_15 can be used for cross-sectional analysis with data from PIAAC-L 2015; it aims at adjusting the figures to the population benchmarks in 2015, at least with regard to some central variables. Bleib_15 is the product of the factors derived from the nonresponse analyses. For longitudinal analyses, this factor should be multiplied with the cross-sectional weight from the previous wave (hrf_14). Accordingly, for longitudinal analysis across multiple waves the cross-sectional weight of the starting wave of interest should be multiplied with the nonresponse weights of the following waves:

- PIAAC 2012 and PIAAC-L 2014:
  For longitudinal analysis of PIAAC 2012 and PIAAC-L 2014, the final full sample weight from PIAAC 2012 (SPFWT0) should be multiplied with the nonresponse weight of PIAAC-L 2014: *SPFWT0 * bleib_14*
- PIAAC 2012 and PIAAC-L 2015:
  For longitudinal analysis of PIAAC 2012 and PIAAC-L 2015, the final full sample weight from PIAAC 2012 (SPFWT0) should be multiplied with the nonresponse weight from PIAAC-L 2014 and the nonresponse weight from PIAAC-L 2015: *SPFWT0 * bleib_14 * bleib_15*
- PIAAC-L 2014 and PIAAC-L 2015:
  For longitudinal analysis of PIAAC-L 2014 and PIAAC-L 2015, the cross-sectional weight from PIAAC-L 2014 (hrf_14) should be multiplied with the nonresponse weight from PIAAC-L 2015: *hrf_14 * bleib_15*
- PIAAC 2012, PIAAC-L 2014 and PIAAC-L 2015:
  For longitudinal analysis of PIAAC 2012, PIAAC-L 2014 and PIAAC-L 2015 the final full sample weight from PIAAC 2012 (SPFWT0) should be multiplied with the nonresponse weights from PIAAC-L 2014 and 2015: *SPFWT0 * bleib_14 * bleib_15*

Please keep in mind that the reference population as described above is limited to a certain age group and excludes people who moved to Germany after 2012. Also, only anchor persons—those who had participated in PIAAC 2012—have a weighting factor. The information provided by other persons in the household can be used as context information in the analyses.

4.2 Reduction of Bias

Table 7 gives an overview of the reduction of bias through weighting in PIAAC 2012, PIAAC-L 2014, and PIAAC-L 2015. In a first step, central sociodemographic indicators were estimated for the PIAAC 2012 cross-sectional sample by applying the PIAAC 2012 final full sample weight (SPFWT0). In a second step, a longitudinal estimation for the same sociodemographic indicators was conducted for the anchor persons' net sample in 2014, multiplying the final full sample weight of PIAAC 2012 with the nonresponse weight of PIAAC-L 2014 (SPFWT0 x bleib_14). In addition, a longitudinal estimation for the anchor persons' net sample of PIAAC-L 2015 was performed, multiplying the final weight of PIAAC 2012 with the nonresponse weights of PIAAC-L from 2014 and 2015 (SPFWT0 x bleib_14 x bleib_15). For PIAAC-L 2015 both, the unweighted results (raw) and the results after weighting are displayed in Table 7. Table 7 shows that bias induced by nonresponse is reduced after weighting for a number of key variables.

Table 7 Reduction of Bias Through Weighting

	PIAAC 2012 (N=5,379)[17]	**PIAAC-L 2014** (N=3,758)	**PIAAC-L 2015** (N=3,263)	
	weighted	weighted	raw	weighted
Gender				
Male	50.5	50.2	48.7	50.1
Female	49.5	49.8	51.3	49.9
Total	100.0	100.0	100.0	100.0
Highest education level including those still in school*				
Low education level	31.6	32.0	23.1	31.7
Middle education level	34.4	33.6	36.5	34.2
High education level	30.7	30.9	36.7	30.7
Still in school	3.3	3.4	3.7	3.5
Total[18]	100.0	100.0	100.0	100.0
Birth cohort				
1946-1961	32.0	32.7	30.8	32.7
1962-1976	34.3	35.1	34.0	35.5
1977-1995	33.7	32.2	35.2	31.8
Total	100.0	100.0	100.0	100.0

*based on information from PIAAC 2012

[17] Eighty-six cases classified as literacy-related nonrespondents are included in the PIAAC 2012 net sample and thus have a weighting factor in 2012. They are also included in the nonresponse adjustment in PIAAC-L. In 2012, only the information on age and gender was collected for them. As information on education as a variable of interest for the analyses of the reduction of bias is missing for these 86 cases, the PIAAC 2012 sample used here consists of N=5,379 (of N=5,465) cases.

[18] Please note: For the sake of clarity values are rounded, thus the cell counts presented here may not add up to exactly 100.0%.

5 Summary and Outlook

This documentation describes the weighting procedure for the second wave of PIAAC-L. Similar to PIAAC-L 2014, the weighting strategy applied here basically follows the approach of the SOEP as a panel study, but also takes into account the weighting procedure in PIAAC 2012. The results are weights that can be used both for longitudinal and for cross-sectional analyses. These weights were delivered for the first time as part of the PIAAC-L data release in December 2016.

Selectivity in PIAAC and PIAAC-L was detected for birth cohorts, in the area of education, as well as other variables. The use of weights is thus recommended for analysis. As mentioned earlier, to account for selectivity between PIAAC-L 2014 and PIAAC-L 2015, for instance, users should multiply the respective weighting factor of PIAAC-L 2014 with the weighting factor bleib_15 (hrf_14 x bleib_15). Using hrf_15 will adjust the figures to the population benchmarks in 2015 and should be chosen for cross-sectional analyses of the data from PIAAC-L 2015.

Nonresponse analyses will also be implemented for the third and last wave of PIAAC-L and user-friendly longitudinal weight will be provided.

References

Bartsch, S., Poschmann, K., & Burkhardt, L. (2017). *Weighting in PIAAC-L 2014.* GESIS Papers 2017|06. GESIS: Cologne.

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse-Bias. A Meta-Analysis. *Public Opinion Quarterly, 72*, 167-189.

Kroh, M. (2014). *Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2012).* SOEP Survey Papers 177, Series D. DIW/SOEP: Berlin.

Kroh, M., Käppner, K., & Kühne, S. (2014). *Sampling, Nonresponse, and Weighting in the 2011 and 2012 Refreshment Samples J and K of the Socio-Economic Panel.* SOEP Survey Papers 260, Series C. DIW/SOEP: Berlin.

Kroh, M., Siegers, R., & Kühne, S. (2015). Gewichtung und Integration von Auffrischungsstichproben am Beispiel des Sozio-oekonomischen Panels (SOEP). In J. Schupp & C. Wolf (Eds.), *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen* (pp. 409-444). Wiesbaden: Springer VS.

Mohadjer, L., Krenzke, T., & Van de Kerckhove, W. (2013). Survey Weighting and Variance Estimation. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 15). Paris: OECD.

Rammstedt, B., Martin, S., Zabal, A., Carstensen, C., & Schupp, J. (2017). The PIAAC Longitudinal Study in Germany – Rationale and Design. *Large-scale Assessments in Education 5*(4).

Saßenroth, D., Kroh, M., & Wagner, G. G. (2013). *Selectivity Processes in and Weights for the Berlin Aging Study II (BASE-II).* SOEPpapers 608. DIW Berlin: Berlin

Spieß, M. (2010). Der Umgang mit fehlenden Werten. In C. Wolf & H. Best (Eds.), *Handbuch der sozial-wissenschaftlichen Datenanalyse* (pp. 117-142). Wiesbaden: VS Verlag für Sozialwissenschaften.

Statistisches Bundesamt (2016). *Ergebnisse des Mikrozensus 2015.* Wiesbaden: Statistisches Bundesamt.

Steinacker, G., & Wolfert, S. (2017). *Durchführung der 2. Erhebungswelle von PIAAC-L. Feldbericht zur Erhebung 2015.* GESIS Papers 2017|04. GESIS: Cologne.

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch, 127*(1), 139-169.

Yamamoto, K., Khorramdel, L., & von Davier, M. (2013). Scaling PIAAC Cognitive Data. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 17). Paris: OECD.

Zabal, A., Martin, S., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I., & Rammstedt, B. (2014). *PIAAC Germany 2012. Technical report.* Münster: Waxmann.

Zabal, A., Martin, S., & Rammstedt, B. (2016). *PIAAC-L Data Collection 2014: Technical Report. Follow-up to PIAAC Germany 2012.* GESIS Papers 2016|17. GESIS: Cologne.

Appendix

Table A1 List of Abbreviations of Independent Variables Used in the Nonresponse Analyses

Variable	Label	Value
Interview Characteristics		
Interviewer change	The interviewer changed from PIAAC-L 2014 to PIAAC-L 2015	1/0
Phone contact	Number of initial contact attempts via phone	1/0
Part. unit nonresp.	Nonresponse of eligible household member in PIAAC-L 2014	1/0
HHmoved	Respondent moved to another address	1/0
Region and Household Characteristics		
Federal state		
Schleswig-Holstein	Schleswig-Holstein	1/0
Hamburg	Hamburg	1/0
Lower Saxony	Lower Saxony	1/0
Bremen / Bremerhaven	Bremen / Bremerhaven	1/0
North Rhine-Westph.	North Rhine-Westphalia	1/0
Hesse	Hesse	1/0
Rhineland-Palatinate	Rhineland-Palatinate	1/0
Baden-Wuertt.	Baden-Wuerttemberg	1/0
Bavaria	Bavaria	1/0
Saarland	Saarland	1/0
Berlin	Berlin	1/0
Brandenburg	Brandenburg	1/0
Saxony	Saxony	1/0
Saxony-Anhalt	Saxony-Anhalt	1/0
Thuringia	Thuringia	1/0
Size of municipality (grouped)		
<20k (rural)	Less than 20.000 inhabitants	1/0
100-500k inh.	100.000 to 500.000 inhabitants	1/0
>500k inh.	More than 500.000 inhabitants	1/0
Housing / building type		
Farm house	Farm house	1/0
Row / duplex house	Row house or duplex (with one dwelling next to the other)	1/0
Building: 3-4 flats	Residential building containing 3 or 4 dwellings	1/0
Building: 5-8 flats	Residential building containing 5 to 8 dwellings	1/0
>8 flats <9stories	Residential building containing 9 or more dwellings (up to 8 stories)	1/0
>9 stories	High-rise building (9 or more stories)	1/0
Building: miss	Missing information on housing / type of building	1/0

Variable	Label	Value
Home owner	Owner of dwelling	1/0
Subtenant	Subtenant of dwelling	1/0
Household size (anchor)		
One person in HH	One person in household	1/0
2 persons in HH	Two persons in household	1/0
3 persons in HH	Three persons in household	1/0
4 persons in HH	Four persons in household	1/0
Income and Employment Characteristics		
Income (quartiles)		
2q. (<=2870 euros)	Monthly net household income less / equal 2870 euros	1/0
3q. (<=4000 euros)	Monthly net household income less / equal 4000 euros	1/0
4q. (>4000 euros)	Monthly net household income more than 4000 euros	1/0
Inc. miss.	Information on monthly net household income is missing	1/0
Labor force participation		
Not in labor force	Not participating in labor force at all	1/0
Occupational position		
Self-empl.	Self-employed (also: Working for a self-employed relative)	1/0
Blue-collar	Blue-collar worker ("Arbeiter"), including those working in agriculture	1/0
Civil serv.	Civil servant (including judges and professional soldiers)	1/0
White-collar	White-collar worker	1/0
Job change	New work / changed position after Dec. 31st 2012	1/0
Family and Partnership Characteristics		
Has partner	Respondent has a steady partner	1/0
Married	Respondent is married	1/0
Divorced	Respondent is divorced	1/0
Family in HH	Family members live in household (incl. spouse)	1/0
Children	Respondent has at least one child	1/0
Education Characheristics		
Education (school degree, grouped)	Education level at the time of the interview in PIAAC-L 2014 (grouped)	
Low education	Low educational level	1/0
High education	High educational level	1/0
Pupil in school	Pupil in school	1/0

Variable	Label	Value
Education parents (school degree, grouped)	Parental education level at the time of the interview in PIAAC-L 2014 (grouped)	
Edu. miss. both. par.	Information on education is missing for both parents	1/0
Low education par.	Both parents have a low education level	1/0
High education par.	Both parents have a high education level	1/0
Edu. mother miss.	Information on education is missing for the mother	1/0
Edu. father miss.	Information on education is missing for the father	1/0
Other Characteristics		
Age (grouped)	Age at the time of the interview in PIAAC-L 2014	
Age 18-24	18-24 years old	1/0
Age 25-34	25-34 years old	1/0
Age 45-54	45-54 years old	1/0
Age 55-68	55-68 years old	1/0
German citizenship	Respondent has German citizenship	1/0
Native speaker: Ger.	Respondents' native language is German	1/0
Disability	Respondent is severely disabled	1/0
Cognitive Skills Characteristics		
Literacy (quartiles)		
Lit 2.q.	Literacy scale score - Competence data 2012 / background data 2012, 2014 – second quartile	1/0
Lit 3.q.	Literacy scale score - Competence data 2012 / background data 2012, 2014 – third quartile	1/0
Lit 4.q.	Literacy scale score - Competence data 2012 / background data 2012, 2014 – fourth quartile	1/0
Numeracy (quartiles)		
Num 2.q.	Numeracy scale score - Competence data 2012 / background data 2012, 2014 – second quartile	1/0
Num 3.q.	Numeracy scale score - Competence data 2012 / background data 2012, 2014 – third quartile	1/0
Num 4.q.	Numeracy scale score - Competence data 2012 / background data 2012, 2014 – fourth quartile	1/0
Problem solving (quartiles)		
PSL q. miss	No plausible values for PSTRE in PIAAC Germany 2012	1/0
PSL 2.q.	PSTRE scale score - Competence data 2012 / background data 2012, 2014 – second quartile	1/0
PSL 3.q.	PSTRE scale score - Competence data 2012 / background data 2012, 2014 – third quartile	1/0
PSL 4.q.	PSTRE scale score - Competence data 2012 / background data 2012, 2014 – fourth quartile	1/0