gesis Leibniz-Institut für Sozialwissenschaften

GESIS *Papers* 2016|17

PIAAC-L Data Collection 2014: Technical Report

Follow-Up to PIAAC Germany 2012

Anouk Zabal, Silke Martin, and Beatrice Rammstedt

GESIS Papers 2016|17

PIAAC-L Data Collection 2014: Technical Report

Follow-Up to PIAAC Germany 2012

Anouk Zabal, Silke Martin, and Beatrice Rammstedt

GESIS – Leibniz-Institut für Sozialwissenschaften 2016

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

PIAAC-L was funded by the Federal Ministry of Education and Research, Berlin (Grant number 01 JP 1301 A, B, C).
For helpful comments on a first draft of this article we thank Simone Bartsch.

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

Contents

1 Introduction

The international PIAAC survey (*Programme for the International Assessment of Adult Competencies*) is part of a large-scale undertaking of the *Organisation for Economic Cooperation and Development* (OECD) to assess and monitor key adult competencies and hereby inform policy decisions. Countries participating in the first cycle of PIAAC conduct a large-scale assessment of literacy, numeracy, and problem solving in technology-rich environments[1] together with the administration of a background questionnaire. This questionnaire covers numerous antecedents of skills as well as economic and social outcomes. National Project Managements for PIAAC are required to adhere to a strict and elaborate set of standards and guidelines in order to ensure a high quality of data and the comparability of the results. The 24 countries participating in the first round of PIAAC (including Germany) carried out their data collection in 2011/2012 (referred to as PIAAC 2012).[2] The international results were made public in October 2013 (OECD, 2013; Rammstedt, 2013) and were taken up by a wide audience. The international Public Use Files are available at the OECD website.[3]

To cater for a very strong interest in the data in the scientific community, Germany published a Scientific Use File (GESIS Data Archive, Data File ZA5845, current version 2.2.0, *doi:10.4232/1.12660*) which included more detailed German PIAAC 2012 data and thus offered more analytical potential than the Public Use Files. At the same time, researchers and policy-makers alike showed interest in addressing more and more elaborate (and, in part, more specifically national) research questions which went beyond the realm of possibilities of the PIAAC data set.

The PIAAC Longitudinal project (PIAAC-L) in Germany arose to at least partly fulfill some of these gaps. In addition, there was a strong interest in increasing the links and synergies between major national surveys. For example, it is quite unclear how the PIAAC proficiency scores for literacy and numeracy relate to the reading and mathematics proficiency scores from the National Educational Panel Survey (NEPS). In terms of obtaining more varied and fine-grained contextual information on the PIAAC respondents (the PIAAC background questionnaire was restricted due to time considerations), the German Socio-Economic Panel (SOEP) offers a well-established concept and elaborate measurement instruments. Thus, PIAAC-L was born as a collaborative project carried out by GESIS – Leibniz Institute for the Social Sciences (lead; GESIS was entrusted with the National Project Management for PIAAC Germany 2012), the Socio-Economic Panel at the German Institute for Economic Research (DIW Berlin), and the LIfBi – Leibniz Institute for Educational Trajectories (LIfBi runs the NEPS). PIAAC-L is funded by the Federal Ministry of Education and Research.

PIAAC-L extends the PIAAC Germany 2012 data through its longitudinal design which comprehends three follow-up waves of data collection in 2014, 2015, and 2016. It further enriches and enhances the contextual information available about PIAAC 2012 respondents by administering a wide variety of questionnaire questions as well as by administering different cognitive instruments, including a re-assessment of literacy and numeracy. As an additional contextual extension, it also includes other adults aged 18 and above living in the same household as the targeted PIAAC Germany 2012 respondents. The PIAAC-L project aims at continuing the high standards of PIAAC and generating comprehensive high-quality data linking PIAAC (with its international comparative context) with nationally-oriented approaches such as the NEPS and SOEP.

---

1 Problem solving in technology-rich environments is an international option and is thus not conducted by all participating countries.

2 The first cycle of PIAAC includes two further rounds of PIAAC with a total of 14 additional countries.

3 http://www.oecd.org/skills/piaac/publicdataandanalysis/#d.en.408927
The data for the Round 2 countries are also available at the OECD website, as will the data for the Round 3 countries.

The first follow-up wave of data collection in 2014 implements SOEP questionnaires and targets PIAAC Germany 2012 respondents—referred to as *anchor persons*—and their adult household members aged 18 and over. The second wave of data collection in 2015 implements (a) a hybrid questionnaire with a block of questions from PIAAC as well as questions from several other surveys, and (b) a cognitive assessment using the PIAAC instruments for literacy and numeracy, and NEPS instruments for reading and mathematics. In the second wave, the interview is administered to anchor persons and, where applicable, their partners living in the same household. The third and final wave of data collection in 2016 is conceptually analogous to the first PIAAC-L data collection, with some new questions included in the SOEP-based questionnaire. In addition, the SOEP cognitive skills battery is administered. The general PIAAC-L design is described in Rammstedt, Martin, Zabal, Carstensen, and Schupp (in press).

Box 1.1 offers a succinct summary of key facts for PIAAC-L in general. Key facts about the first wave of data collection in 2014 are summarized in Box 1.2.

Box 1.1: Key Facts: PIAAC-Longitudinal (PIAAC-L), Germany

- National longitudinal follow-up of PIAAC Germany 2012 respondents
- Three waves of data collection: 2014, 2015, 2016
- Main objective is to extend and enhance analytical potential of German PIAAC 2012 data
- Cooperative undertaking between three major Leibniz institutes and surveys:
  - → GESIS – Leibniz Institute for the Social Sciences [lead]
  - → Socio-Economic Panel (SOEP) at German Institute for Economic Research (DIW Berlin)
  - → Leibniz Institute for Educational Trajectories (LIfBi)
- Survey organisation: TNS Infratest
- Funded by the Federal Ministry of Education and Research

Box 1.2: Key Facts: The Data Collection 2014 (PIAAC-L Wave 1)

- Instruments (adapted from SOEP core instruments):
  → Household questionnaire
    - Living situation, conditions, and costs
    - Household income and benefits, wealth
    - Children and other household members

  → Person questionnaire
    - Background information, family, childhood
    - Biographical calendar
    - Formal education (general and vocational education), continuing professional education
    - Work status, work situation and history
    - Income and benefits
    - Health, time use, leisure activities, attitudes, personality, opinions, satisfaction
- Interview administration: CAPI (computer-assisted personal interview), usually administered in the respondent's home
- Interview language: German
- Target persons:
  → Anchor persons (participated in PIAAC Germany 2012 and could be contacted for PIAAC-L)
  → All household members in the anchor persons' household aged 18 and over
- Data collection period: February 26 to August 14 2014
- Number of interviewers: 138 (116 with PIAAC experience)
- Interview duration (on average):
  → Household protocol: 3.5 minutes
  → Household questionnaire: 16 minutes
  → Person questionnaire: 44 minutes
- Gross sample size (anchor persons): 5 225
- Realized sample size:
  → Anchor persons: 3 758
  → Household members: 2 473
- Achieved response rate: 72%
- Data: accessible for scientific purposes as scientific use files (ZA5989) from GESIS Data Archive / Research Data Centre PIAAC (FDZ PIAAC)[4]

[4] Version accessible on 1.12.2016: GESIS – Leibniz Institute for the Social Sciences, German Socio-Economic Panel (SOEP) at DIW Berlin & LIfBi – Leibniz Institute for Educational Trajectories (2016): PIAAC-Longitudinal (PIAAC-L), Germany. GESIS Data Archive, Cologne. ZA5989 Data file Version 1.1.0, *doi:10.4232/1.12576*

This technical report describes the implementation of the first wave of data collection for PIAAC-L. After the brief overview of PIAAC-L and the first data collection in 2014 offered in this first chapter, we describe the transition from PIAAC to PIAAC-L in Chapter 2. Chapter 3 provides information on the survey instruments. Chapter 4 focusses on fieldwork and fieldwork results. A summary of data management activities and the data products is then given in Chapter 5.

2 The Transition From PIAAC to PIAAC-L

PIAAC was designed as a cross-sectional study. However, as the PIAAC-L coordinator already had plans in place to convert PIAAC 2012 Germany into a longitudinal study, the German PIAAC respondents were asked at the end of the PIAAC interview whether they could be re-contacted for a possible follow-up survey. Ninety-six percent of the PIAAC 2012 respondents in Germany could be re-approached.

2.1 Panel Maintenance

As a longitudinal (panel) survey, non-response in PIAAC-L can occur at each wave of data collection resulting in sample attrition, meaning that the effect of loss of respondents (sampling units) is generally cumulative over the waves of data collection. To counteract this it is important to implement adequate measures of panel maintenance. With a view to the rather lengthy period of time between fieldwork in PIAAC Germany in 2011/2012 and the first PIAAC-L data collection scheduled for February 2014, a panel maintenance activity was carried out at the end of 2013. All PIAAC Germany 2012 respondents (for which a further contact was admissible) were sent a PIAAC information brochure (Rammstedt et al., 2013a)[5] summarizing the German national results and the international comparisons by postal mail. The enclosed cover letter thanked the respondents for their valuable contribution to the success of the German PIAAC survey, and announced that a follow-up of PIAAC was underway and they would be contacted for this project in due time. This panel maintenance measure served various purposes: For one, it engaged the PIAAC-L target persons, reminding them of their participation in PIAAC, pointing out that first survey results had been disseminated in the media in the past weeks, and providing them as respondents personally with a summary of the PIAAC results. Furthermore, it introduced the new national PIAAC follow-up (PIAAC-L), emphasizing that their personal participation was crucial to this follow-up survey since only PIAAC 2012 respondents were eligible for PIAAC-L. Lastly, this contact was used to check the available address information. Where required, the appropriate registry offices were contacted and asked to provide updated address information.

2.2 Terms of Agreement for the PIAAC-L Project

The first wave of data collection in PIAAC-L was especially crucial as the potential *anchor persons,* i.e. the former PIAAC 2012 respondents, were introduced to the (new) project PIAAC-L and the institutes collaborating in this project. Recruitment at this stage was defined to cover the entire PIAAC-L project, i.e. all three waves of data collection. Thus, the anchor persons were asked to enter a new agreement regarding their participation in a follow-up survey to PIAAC which was nevertheless a "new" survey. Some constants and differences between PIAAC Germany 2012 and PIAAC-L are notable in this context:

(1) The survey organisation contracted for PIAAC (TNS Infratest) also won the bidding process for PIAAC-L. Thus, the target persons were familiar with the PIAAC-L fieldwork agency. In addition, it was not necessary to request permission to transmit address data to another (new) fieldwork institution.

[5] The English version of this brochure (Rammstedt et al., 2013b) is available at http://www.gesis.org/fileadmin/piaac/Downloadbereich/PIAAC_Zusammenfassung_engl.pdf.

(2) The project management for PIAAC-L is the collaborative responsibility of three institutions, and no longer the sole responsibility of GESIS, as was the case for PIAAC Germany. However, since GESIS has the lead coordinating role for PIAAC-L, it remains prominent and familiar (trustworthy) to target respondents.

(3) PIAAC-L is a national project, and no longer has the shine of internationality.

(4) PIAAC-L addressed not only the PIAAC respondents themselves, but also other resident members of their households ages 18 and above.

All persons interviewed for PIAAC-L, i.e. both the anchor persons and household members, were required to consent to having their data linked with (1) the anchor person's PIAAC 2012 data, (2) the data of all other members of their household interviewed for PIAAC-L, including the data of the anchor person and the other household members from the two subsequent PIAAC-L waves in 2015 and 2016.

2.3 The PIAAC-L Sample

As a follow-up to the cross-sectional survey PIAAC 2012, PIAAC-L did not select a new probability-based sample, but is instead based on the sample of PIAAC Germany 2012 respondents (N=5 465). The PIAAC-L gross sample only includes PIAAC 2012 respondents that could be legally re-contacted for a follow-up survey. At the end of the PIAAC Germany 2012 interview, respondents were asked whether they could envision participating in a prospective PIAAC-related panel survey. Respondents who said they would not consider participating in such a panel survey were removed from the sample (about 2%).[6] Respondents who did not negate this question could be re-approached at a later date for a follow-up survey. Thus, the initial primary gross sample consisted of 5 225 PIAAC Germany 2012 respondents (see Table 2.1). Please note that as also indicated by Zabal et al. (2014), the PIAAC Germany 2012 survey achieved 5 465 completed cases as defined by the international PIAAC standards. This included 86 so-called literacy-related nonrespondents (who were exempt from a full interview) and 59 breakoffs that were approved as completed cases by the international PIAAC Consortium. Both these groups were excluded from the PIAAC-L gross sample.

The *anchor persons* build the backbone of the PIAAC-L sample. However, PIAAC-L extended the target population to also include adult household members in the anchor person's household (the operational definition for wave 1 was household members born in 1996 or earlier). While this extension to the household provides further context for the PIAAC-L anchor persons, it is important to understand that it does not correspond to a genuine household sample – the primary PIAAC-L sample remains in essence a sample of persons. Participation of the anchor persons is a conditio sine qua non in every wave of data collection in PIAAC-L – other household members are only included in the gross sample on the condition that the corresponding anchor person agrees to participate (or continue to participate) in PIAAC-L.

[6] This refers to the percentage of consenters of those who received the consent question; literacy-related non-respondents and breakoffs were not administered the consent question.

Table 2.1. From the PIAAC 2012 Net Sample to the Initial PIAAC-L Gross Sample

Completed cases (net sample) PIAAC Germany 2012	**5 465**
Literacy-related nonrespondents	86
Breakoffs to the cognitive assessment	59
Respondents who did not consent to a re-contact	95
Initial gross sample PIAAC-L 2014 (anchor persons)	**5 225**

The target population in PIAAC consisted of non-institutionalized adults between 16 and 65 years of age, residing in the country during data collection, irrespective of their nationality, residential status or language skills (Mohadjer, Krenzke, & Van de Kerckhove, 2013a; OECD, 2010). All countries participating in PIAAC were required to realize a probability-based sample representative of this target population. For PIAAC 2012 in Germany, a registry-based, two-stage stratified and clustered sampling design was utilized: In the first stage, a *probability proportionate to size* design was used to select a stratified random sample of municipalities. In the second stage, a systematic random sample of persons was selected from the registries of the municipalities that were selected in the first stage. For further details on the sampling design and selection for PIAAC 2012 Germany see the national technical report for PIAAC Germany 2012 (Zabal et al., 2014).

The German PIAAC 2012 sample was distributed over 277 communities with 320 sample points. Due to mobility and relocation of PIAAC-L respondents over the years after their participation in PIAAC (2011/2012), there were 565 communities in the first wave of PIAAC-L (2014).

3 Instruments

The first wave of data collection was designed with the intention to bring the PIAAC and the SOEP survey closer together. The idea was to adopt an approach closely based on the SOEP and by doing so to enrich the background information available on the anchor persons. This first wave of data collection for PIAAC-L therefore basically implemented core SOEP instruments which were slightly modified to suit the PIAAC-L purposes.

There were three instrument components, all of which were administered by the interviewer: the household protocol, the household questionnaire, and the person questionnaire. The household protocol necessarily needed to be completed first, as information from this protocol was required to determine the target persons in the household and to activate the corresponding interviews. In PIAAC-L, the term household always refers to the *household of the anchor person*. One household questionnaire was required to be completed per household. Person interviews were to be attempted with all household members aged 18 and over. A completed case was defined as a case with a completed household protocol, a completed household questionnaire, and a completed person interview with the anchor person.[7] The goal was to administer also the person questionnaire to all other selected household members, but their participation was not decisive.

Before starting with the person interview, all respondents were asked for their consent to having their data linked with that of the other respondents in their household and for all waves of PIAAC and PIAAC-L data collection (as mentioned in Section 2.2). In compliance with data and privacy protection requirements, PIAAC-L respondents were handed out an information sheet summarizing all the relevant information in order to obtain their informed permission. This information was read out loud by the interviewer, and the respondent's response was registered in CAPI. Respondents who did not give their permission for this data linkage were excluded from PIAAC-L. Less than 0.5%[8] of the respondents that started the interview refused consent for the data linkage (and thus terminated their participation in the PIAAC-L project).

3.1 Household Protocol

A household protocol was administered to record the general household composition and household structure. It was used to determine which resident household members were to be recruited for the PIAAC-L interview (year of birth 1996 or earlier), i.e. the additional target persons beyond the (mandatory) anchor person. Although it was not mandatory that information for the household protocol be provided by the anchor person, in view of the completed case definition for which a completed anchor person interview is a prerequisite, it was recommended to interviewers that they try to complete both the household protocol and the household interview with the anchor person.

Concretely, the electronic household protocol used in the 2014 PIAAC-L data collection listed all persons living in the anchor person's household, and collected the following information for each registered household member:

[7] After data collection, it was established that there were 21 cases with completed household protocols, completed anchor person interviews, but no household interview. An a posteriori decision was made to also accept these as completed cases for this first wave of data collection.

[8] This refers to all PIAAC-L respondents, i.e. both anchor persons as well as participating household members.

- First name
- Year of birth
- Gender
- Relationship to PIAAC anchor person[9]
- Date at which person came to live in household
- Temporary or permanent absences
- Additional residences

The documentation of contact results was integrated into the household protocol, so that final disposition codes were recorded at the "household" level. Thus, anchor person refusals—which directly eliminated the entire household from the project—were also recorded at this level. The contact documentation included type of contact (mode of visit), date and time of contact, and contact result. Some questions about the neighbourhood and dwelling characteristics were also included in the household protocol. The administration of the household protocol took three to four minutes. The household and person interviews were technically activated on completion of the household protocol.

3.2 Household Questionnaire

The content of the household questionnaire implemented in the first wave of PIAAC-L was based on the 2014 SOEP core household questionnaire (TNS Infratest Sozialforschung, 2014a) and was administered as CAPI with showcards for certain questions. In the SOEP, the head of the household is asked to complete this questionnaire as she/he usually has more comprehensive knowledge about household-level information. Contrary to the SOEP, and as a consequence of the anchor person concept followed in PIAAC-L, the household questionnaire was administered to the anchor person if they felt competent to answer the questions.[10] Otherwise, interviewers were instructed to identify and recruit the household member best suited to respond to this questionnaire.

A documentation of the household questionnaire for PIAAC-L 2014 can be found at the GESIS Archive website (ZA5989_fb_hh.pdf).[11] The documentation reflects the instrument as it was administered in the field and is therefore in German. An English translation of the questions is provided in the codebook (ZA5989_cod_Household_14.pdf).[12]

The household questionnaire covers the following topics:

- **Living situation, living conditions and costs**: type of dwelling, characteristics of living space, facilities (including internet and phone connections), condition of building, neighbourhood characteristics, ownership/tenancy, rent, mortgage, maintenance and additional costs
- **Household income and benefits, wealth**: household income and detailed sources, government aids and subsidies, savings, loans
- **Children and other household members**: year of birth and sex of children living in the household, childcare, school attendance, extra-curricular activities, members of household requiring help, help providers/carers, various costs related to children and care of household members

The administration duration for the household questionnaire was on average 16 minutes.

---

[9] Due to the anchor person concept followed in PIAAC-L, the household protocol collected information on the relationship of each household member to the anchor person (and not the head of household, as is the case in the SOEP survey).

[10] 92% of the household interviews were carried out with anchor persons.

[11] https://dbk.gesis.org/dbksearch/download.asp?db=D&id=59049

[12] https://dbk.gesis.org/dbksearch/download.asp?db=D&id=59045

3.3 Person Questionnaire

The PIAAC-L 2014 person questionnaire was administered to anchor persons and to all additional adult household members who agreed to participate in PIAAC-L. The person questionnaire covers the following topics:[13]

- **Background information**: year of birth, country of birth, nationality (incl. second citizenship, German citizenship since birth, intent to apply for German citizenship), immigration, attachment to country of origin, identification as German, living and household situation, childhood (e.g. childhood grades obtained in last school report card or home situation in childhood), current time use, life events (after 31.12.2012)
- **Family**: marital/civil status, partnership (information regarding current and previous relationships, cohabitation, marriage/separations), children (e.g. year of birth, sex, place of residence), parental information (birth, death, nationality, education, occupation, religion), siblings (e.g. year of birth, sex, type), payments to family members
- **Biographical calendar** starting at age 15 and up to a maximum of age 65 with respect to education and employment
- **Education**, i.e. formal education (general and vocational education) and continuing professional education: location of school and professional education, years of schooling, qualifications (incl. year qualification was obtained), recognition of foreign qualifications, current education, current vocational education or further education (incl. number of training programs attended, number of days, training providers, training costs, barriers to participation), educational plans
- **Work status, situation, and history**: current/first/last occupation/occupational status, voluntary/military service, current/first/last industry, current job characteristics (sector, company size, contract conditions, working hours), current employment status, unemployment, work changes, side jobs
- **Income and benefits** (current and last year): wages, bonuses, benefits, income sources and gross income per source
- **Health, attitudes, personality**: current health (e.g. SOEP SF-12 short version, disability, smoking, medical appointments, hospital spells), satisfaction with different areas, general risk propensity, big five, trust, reciprocity, grit, political inclination, voting behaviour, attitude towards lifelong learning, feelings, worries

The person questionnaire for this first wave of data collection was based on two SOEP core questionnaires: the *Biography Questionnaire* ("Lebenslauf") for 2014 (TNS Infratest Sozialforschung, 2014b), which contains a wide variety of biographical questions typically administered only once to SOEP respondents, and the *Individual Questionnaire* ("Personenfragebogen") for 2014 (TNS Infratest Sozialforschung, 2014c). Due to time constraints and the fact that some of the information had already been assessed as a part of the PIAAC interview, the PIAAC-L person questionnaire did not contain all questions originally included in these two questionnaires; an overview of the questions which were excluded can be found in the Annex. In addition to these two questionnaire sources, a number of questions on non-cognitive skills were added to the PIAAC-L person questionnaire for wave 1 in order to extend information on potential factors related to the basic cognitive skills measured in PIAAC. Concretely, the following additions were undertaken: (1) the SOEP *Big Five* item battery,[14] (2) the SOEP *Locus of Control* item set,[15] (3) SOEP items measuring *Reciprocity* and other attitudes,[16]

[13] The list of topics and sub-topics is ordered according to topic and not order in questionnaire.

[14] ZA5989_fb_persbio.pdf: Question set L71, variables pego01_14-pego16_14

[15] ZA5989_fb_persbio.pdf: Question set L188, variables pzu01_14-pzu09_14

[16] ZA5989_fb_persbio.pdf: Question set L187, variables pmas01_14-pmas11_14

and (4) a small set of items that had been included in the PIAAC field test (cp. OECD, 2011) but which were subsequently excluded from the PIAAC main survey due to length constraints. Of the latter, five items[17] originated from the *Grit*[18] scale (Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Quinn, 2009): Four items tapped the sub-construct *Perseverance of Effort* and one item measured the sub-construct *Consistency of Interests*. In addition, another item tapping perseverance was included.[19] The last item taken from the PIAAC field test had been newly developed for PIAAC and was intended to yield a measure for surface learning.[20]

A documentation of the final person questionnaire for PIAAC-L 2014 and the corresponding codebook can be found at the above-mentioned GESIS Archive website (ZA5989_fb_persbio.pdf; ZA5989_cod_Persons_14.pdf).[21]

The person questionnaire was administered as CAPI. For specific questions the respondent was handed a set of showcards to be referred to. On average, the administration time of the person questionnaire was 44 minutes.

[17] ZA5989_fb_persbio.pdf: Question set L73, variables piq01_14-05_14

[18] Grit is regarded by Duckworth and Quinn (2009) as a perseverance of efforts and commitment towards more long-term goals.

[19] ZA5989_fb_persbio.pdf: Question set L73, variable piq06_14

[20] ZA5989_fb_persbio.pdf: Question set L73, variable piq07_14

[21] Questionnaire: https://dbk.gesis.org/dbksearch/download.asp?db=E&id=59050;
codebook: https://dbk.gesis.org/dbksearch/download.asp?db=E&id=59046

4 Fieldwork

After a competitive bidding process, the survey organisation TNS Infratest was subcontracted by GESIS, as the coordinator of the PIAAC-L Consortium, to collect the data for all three waves of PIAAC-L. Data collection for the first wave of PIAAC-L took place between February 26, 2014 and August 14, 2014.

4.1 Staff

The staff for PIAAC-L at the survey organisation consisted of two field directors, seven supervisors, and 138 interviewers. Both field directors had been responsible for the PIAAC 2012 survey at the fieldwork agency. The majority of the interviewers (116) were PIAAC 2012 interviewers, and all were experienced with the SOEP survey and as such all were high-performing interviewers experienced in CAPI administration and familiar with the specific questionnaires used in PIAAC-L wave 1. The majority of interviewers were over 50 years of age (11% were under 50 years of age, 27% over 70 years of age) and had worked for the survey organisation for longer periods of time (only 18% had worked for the survey organisation less than five years, 59.5% five to fifteen years, 22.5% over 15 years). Fifty-four percent of the interviewers were male. Forty-eight percent had a middle level of education and 38% a high level of education. There was no interviewer attrition, although there were very few cases in which interviewers were not available for a part of the fieldwork period, for example due to sickness.

4.2 Interviewer Training and Briefing

As previously indicated, the first wave of PIAAC-L data collection is very closely based on the SOEP core survey. All interviewers were experienced with the administration of the SOEP survey, and more specifically the SOEP questionnaire content used in the PIAAC-L 2014 data collection. Thus – unlike the PIAAC survey – no extensive systematic training was required. However, a personal face-to-face training was regarded as useful for interviewers without PIAAC experience. Beyond the PIAAC-L specific training contents, the aim was to introduce the non-PIAAC interviewers to the PIAAC survey and provide them with the appropriate background. This was important to ensure they were aptly prepared to contact the PIAAC 2012 respondents and gain their cooperation for PIAAC-L.

Interviewer trainings were carried out in-person for 48 interviewers. All 22 interviewers without PIAAC experience attended the training, the remaining PIAAC interviewers were primarily invited so they could share their experience with the non-PIAAC interviewers. The half-day training covered the following topics: (1) review of PIAAC 2012, (2) introduction to the PIAAC-L project and its aims, (3) the PIAAC-L design, (4) survey material, fieldwork phases, and survey procedures, (5) obtaining respondent cooperation for PIAAC-L, (6) data privacy issues, most notably obtaining the data linkage consent, (7) quality control, and (8) differences between PIAAC-L and the SOEP. All other interviewers were provided with training via a WebEx conference, using a somewhat condensed and edited version of the face-to-face training material.

Interviewers were also provided with an interviewer manual and, as in PIAAC, an interviewer booklet. The manual included detailed design and procedural information, including:

- Key facts about PIAAC 2012
- Key facts about PIAAC-L: overall design, the three research institutes cooperating for PIAAC-L
- Sample: target population, anchor person concept, household extension, definition of completed case
- Obtaining consent to data linkage: background and implementation
- Incentives
- CAPI instruments and accompanying survey material
- Survey procedures
- Case documentation and disposition codes
- How to conduct practice interviews
- A register listing and explaining crucial terms used in the questionnaires (adopted from the SOEP survey)

In addition, key information for the administration of the PIAAC-L interviews was summarized in a compact interviewer booklet. A condensed version of the interviewer training presentations was also distributed to interviewers. Furthermore, interviewers received written instructions and briefings (focusing on administrative issues) as required during fieldwork.

4.3 Interviewer Remuneration

The interviewer payment scheme for PIAAC-L was aligned with the one adopted for PIAAC. In PIAAC Germany 2012 the interviewer remuneration included: (1) an above-average rate per completed interview, reflecting the length, complexity, and prominence of the survey, (2) an add-on for completed interviews in large municipalities to compensate for the generally lower cooperation rates, (3) an additional hourly component for long interviews to ensure that respondents would be given the time they required for the assessment, (4) reimbursement of all travel expenses, and (5) day rates for specific situations (cp. Zabal et al., 2014). The payment scheme was very similar for PIAAC-L, with only two slight modifications. First, the rate per completed interview was slightly reduced for PIAAC-L (which is also less complex and salient than PIAAC). However, the PIAAC-L rate remained more attractive than the rates offered for comparable national surveys. Second, no additional hourly rate was payed for long interviews since this first wave of PIAAC-L did not include an assessment component.

4.4 Addressing Respondents

The first PIAAC-L wave of data collection in 2014 recruited target anchor persons' cooperation for the PIAAC-L project, and as such, success in gaining respondent cooperation in this wave was crucial. Various efforts were undertaken to support interviewers in engaging respondents for PIAAC-L.

Two types of incentives were offered to the respondents. First, an unconditional non-monetary incentive was sent together with an advance letter. This consisted of four postage stamps (at 60 cents each), enclosed in a card bearing the PIAAC-L logo. Second, respondents received a conditional monetary incentive according to the following scheme:

- Upon completion of the household protocol, the household questionnaire, and the interview with the anchor person (person questionnaire): 25 euros
- Each additional completed person interview: 10 euros

The monetary incentives are lower than those offered for PIAAC (Zabal et al., 2014), but higher than those usually provided in the SOEP survey (Glemser, Huber, & Bohlender, 2016).

The target persons (anchor persons) were sent a personalized advance letter introducing PIAAC-L and announcing the visit of an interviewer (introduced by name). The advance letter was accompanied by an information sheet, a confidentiality/data privacy statement, and the above-mentioned postal stamps (unconditional non-monetary incentive). The information sheet provided more details about the PIAAC-L survey and the four institutes responsible for the survey (GESIS, DIW, and LIfBi as the project consortium, as well as TNS Infratest as the survey organisation). It also elaborated on important design features, such as the three waves of data collection and the inclusion of household members. Furthermore, it emphasized why the target persons' participation is especially important while clarifying that participation is absolutely voluntary.

An additional contact letter was prepared for the re-issue phase in which non-contacts were re-worked and refusal conversion efforts were undertaken. This additional contact letter was also accompanied by the information sheet and the data privacy statement. All material intended for the respondent (advance letters, information sheet, data privacy statement, stamps, and callback cards) bore the PIAAC-L logo to increase familiarity with the PIAAC-L project. The PIAAC-L logo retains crucial aspects of the PIAAC logo, so the association between the two surveys is nicely visualized.

TNS Infratest also offered an information hotline for respondents that operated during the usual office hours.[22]

4.5 Fieldwork Procedures and Monitoring

Fieldwork basically consisted of two phases. In the first and main fieldwork phase, all cases (anchor person names and addresses) were simultaneously released at the beginning of fieldwork. As far as possible, PIAAC interviewers were allocated cases they had already worked in PIAAC.[23] Interviewers worked their cases especially intensely during the first weeks of this phase. In the second phase, the following cases were re-issued:

- Non-contacts
- Respondents with a final disposition indicating they were not available during the main fieldwork phase
- Soft refusals: As prescribed by German legislation, only a subset of the refusals – the so-called "soft refusals" – can be re-approached; to determine whether a refusal could be re-issued, all refusal dispositions and the accompanying open entry information as well as information from the respondent hotline were carefully screened
- Respondents for which new addresses were tracked through an address search via the registry offices

[22] Of the approximately 120 calls registered, somewhat over one third were respondent refusals. The rest of the calls revolved around passing on changed address information, providing a phone number or similar information relevant to the interviewer.

[23] On average, PIAAC interviewers were assigned somewhat more cases than interviewers without PIAAC experience.

About 65%[24] of the non-interviews (with anchor target persons) were re-issued in the second fieldwork phase. The re-issue phase was split up in two sub-re-issue periods. In some cases, however, the differentiation in the various working periods was not clear-cut.

Refusal conversion efforts were intensified in this latter phase and focused on gaining the cooperation of anchor persons as obtaining panel cooperation from the anchor persons (achieving as many completed cases as possible) had the highest priority. Although achieving maximal household completeness was also a fieldwork objective, this was not pursued aggressively. Interviewers were instructed to be sensitive in their recruitment of household members and weigh whether their attempts to obtain respondent cooperation for other household members could backfire and cause the anchor person to be annoyed and backtrack from their commitment to the PIAAC-L project. In this second fieldwork phase, interviewers were required to ensure that all adult household members had been addressed and their contact results documented.

Intensive efforts were undertaken throughout fieldwork to track movers and obtain current, valid addresses in order to re-contact as many of the original PIAAC 2012 respondents as possible. More detailed information on fieldwork can be found in the documentation produced by the survey organisation (Steinacker, Schmidt, Wolfert, & Schneekloth, 2016).

Contacting Procedures

Contacting instructions specified that interviewers were to contact the anchor persons shortly after these had been notified of the PIAAC-L project and the upcoming interviewer visit per advance letter. Interviewers left the PIAAC-L contact cards behind if the target person was not at home. At least four in-person contact attempts were required before a non-contact could be coded. Interviewers were encouraged to realize an efficient schedule of contact attempts and spread these out over different weekdays and time of day. Contact history information and final contact results at the level of the anchor person ("household") were documented electronically (cp. Section 3.1). A paper address/contact form was additionally available and used for administrative and invoice purposes.

Quality Control of Fieldwork

The survey organisation monitored the interviewers' performance closely throughout fieldwork. Fieldwork progress reports were regularly sent to and discussed with GESIS. Preliminary data from the field were also provided to the PIAAC-L Consortium at various times during fieldwork.

Interviewer validation encompassed the following three components:

- 100% validation of all complete interviews via a short questionnaire that was sent by postal mail to all anchor persons that had participated in the first wave of PIAAC-L[25]
- A consistency check between the registry data (which was obtained for PIAAC Germany 2012) and the PIAAC-L 2014 interview data with respect to gender and year of birth
- Check of interview duration with a special view to very short interviews

The validation questionnaire asked for feedback on: (1) whether the respondent (anchor person) had been recently interviewed for the PIAAC-L project, (2) whether the interview had been administered with a laptop or a paper questionnaire, (3) respondent birth year and gender, (4) how many other adult household members had also been interviewed, and (5) the duration of the respondent's interview. The validation questionnaire was sent as soon as the status of the anchor person's household was regarded as complete (i.e. home office had a household protocol, a household questionnaire, the anchor person interview data, and either a final non-interview disposition for all

[24] 1 057 of the 1 622 non-interviews.

[25] There was no validation of the interviews with the other household members.

other adult household members or all the corresponding interview data). Response to the validation questionnaire was facilitated by keeping the questionnaire short and including a self-addressed envelope in the validation mail.

Validation was an ongoing, continuous process during fieldwork. The following validation results were automatically flagged for further scrutiny and inquiry:

- More than one year discrepancy between year of birth given in the PIAAC-L 2014 interview or information from validation questionnaire and PIAAC Germany 2012 registry data
- Discrepancies with regard to gender information
- Validation questionnaire indicates that no interview was carried out with anchor person
- Validation questionnaire indicates that interview was not administered face-to-face or validation questionnaire indicates interviewer did not use a laptop for interview administration
- Actual interview duration of less than 20 minutes
- Interview duration as given in validation questionnaire is less than 15 minutes

Any interview that was flagged during the validation process was subsequently individually checked. These follow-ups took all available information into account (all questionnaires that were administered in the anchor person's household, data from PIAAC Germany 2012, validation questionnaire data). Furthermore, the specific interviewer's entire work was considered at this time and checked for conspicuous patterns. Any issues requiring further clarification were addressed by contacting the respondent by phone; this was necessary for only very few cases. If contact remained unsuccessful after various attempts, the interviewer was contacted and a clarification elicited.

Over 60% of the validation questionnaires were completed and sent back by the anchor respondents. 130 anchor person interviews were flagged for further examination during the validation process. The overwhelming majority of the flagged validation results were resolved (e.g. it was determined that the validation questionnaire had been filled out by another household member that had participated in PIAAC-L and not the anchor person, data entry and data management errors were identified, it was clarified that the anchor person interpreted the show card booklet as a paper booklet but the interview had been administered as CAPI). The validation results were reviewed together with GESIS. No falsifications were identified. Two cases were identified in which the interviewer had unintentionally carried out the interview with the wrong person (a case of mistaken identity due to father and son living in the same household). Some data cleaning needs were identified, such as the need to delete interview duplicates due to technical breakoffs and re-administration of the interview.

Finalizing Disposition Codes

At the end of fieldwork, any remaining temporary dispositions were addressed and final disposition codes were determined and finalized for each case. If a case had been worked in more than one fieldwork phase, the final results were reconciled. This was especially important for the final coding of the non-contact disposition: The final disposition was only coded as a non-contact if no contact was made whatsoever with the household during the entire fieldwork (and not, for example, if non-contact was coded only for the last fieldwork phase).[26]

4.6 Fieldwork Results

The fieldwork results refer to the final data after validation wrap-up and data cleaning. They are also based on the completed case definition – a case with a completed household protocol, a completed

[26] This corresponds to the procedures implemented in PIAAC Germany 2012.

household questionnaire, and a completed person interview with the anchor person – although, as previously mentioned, an ex post decision was made to include the few cases for which no household interview was available, but for which there were a valid household protocol and a completed anchor person interview.

The net sample for this first wave of data collection in the project PIAAC-L consists of 3 758 completed interviews with the anchor persons (primary net sample) and 2 473 completed interviews with the household members, yielding a total of 6 231 interviews. There were no breakoffs in this first wave of PIAAC-L. On average, 1.7 persons were interviewed per anchor person household.

The following data are available for the 3 758 completed cases (this corresponds to the anchor person household level):

- 3 758 household protocols
- 3 737 household interviews
- 6 231 person interviews

The final disposition codes for the anchor persons based on the primary gross sample of 5 225 persons are shown in Table 4.1.

Table 4.1. Final Disposition Codes for Anchor Persons

Final disposition code		***n***	%
1	Anchor person unavailable during fieldwork period	142	2.7
2	Anchor person permanently ill or incapable	46	0.9
3	Anchor person institutionalized	3	0.1
4	Language problem, German proficiency insufficient	4	0.1
5	Refusal anchor person	846	16.2
6	Anchor person moved outside of country	31	0.6
7	Anchor person moved, new address unknown	121	2.3
8	Anchor person moved, new address known	10	0.2
9	Anchor person unknown at given address	8	0.2
10	Invalid address	33	0.6
11	Not a residential address, dwelling not inhabited	2	0.0
12	Death	6	0.1
13	Other reasons or unusual circumstances	4	0.1
14	Non-contact	170	3.3
15	Someone is home, but door is not opened	6	0.1
16	Contact established, no final disposition	23	0.4
21	Completed case: household interview and anchor person interview	3 737	71.5
22	Completed case: anchor person interview, no household interview	21	0.4
23	Interview not usable	12	0.2
Total		5 225	100.0

Approximately 19% of the anchor persons lived in single households, 34% in two-person households, and 40% in three-to-four-person households. There was a total of 4 214 adult household members, born in 1996 or earlier (or without information on the year of birth),[27] living in the households of the participating anchor persons that were considered eligible as PIAAC-L household participants—this corresponds to the gross sample for other adult household members. The final disposition codes for the adult household members can be found in Table 4.2.

Table 4.2. Final Disposition Codes for Adult Household Members

Final disposition code		***n***	%
1	Unavailable during fieldwork period	98	2.3
2	Permanently ill or incapable	35	0.8
4	Language problem, German proficiency insufficient	37	0.9
5	Refusal	1 082	25.7
13	Other reasons or unusual circumstances	98	2.3
14	Non-contact	13	0.3
15	Someone is at home, but door is not opened	1	0.0
17	Nonresponse without final interviewer documentation	377	8.9
20	Interview	2 473	58.7
Total		4 214	100.0

As illustrated in Table 4.3, more than half of the eligible adult household members are spouses or partners of the anchor person. Parents or children of the anchor person make up over 35% of the other household members.

[27] For 98% of these cases a final disposition was assigned by the interviewer.

Table 4.3. Relationship of Eligible Adult Household Member to Anchor Person

Relationship to anchor person	*n*	%
1 Son/daughter (biological, adopted, stepchild)	661	15.7
2 Spouse/partner	2 371	56.3
3 Father/mother (biological, adoptive, stepparents)	835	19.8
4 Sibling (biological, stepsibling)	249	5.9
5 Others	89	2.1
6 Relationship not stated	9	0.2
Total	4 214	100.0

The average interview length per average participating household was approximately 93 minutes (Steinacker et al., 2016).

Response Rates

The simple fieldwork response rate based on the primary gross sample of 5 225 is 72%.[28,29,30] The participation rate for the other adult household members is 59%.[31] Two things need to be considered when evaluating these rates: (1) Although PIAAC-L explicitly targets all adult members of the anchor person household, the original sample was a *person sample* (PIAAC Germany 2012 respondent, i.e. the anchor person). Thus, the extension to the household realized in this first wave of PIAAC-L differs from a genuine household-based study which – in the case of the SOEP – usually addresses the head of household and the participation of the entire household from the very beginning.[32] (2) The focus in PIAAC-L is clearly on obtaining anchor person respondent cooperation and recruiting other adult household members was not supposed to endanger the anchor person cooperation (i.e. anchor person household cooperation). Given the transition from a person sample to a household approach, but combined with the clear anchor person concept followed in PIAAC-L, the participation of household members is expected to be lower than in genuine household-based studies (and the participation rates are not directly comparable due to the difference in approach and the completed case definition).

[28] 3 758/5 225

[29] Given the very specific and unique calculation of PIAAC response rates (cf. Mohadjer, Krenzke, & Van de Kerckhove, 2013b; Zabal et al., 2014), which is no longer directly applicable to the PIAAC-L design, and the fact that there is no standard and straightforward application of the AAPOR response rates to registry-based samples, we report the simple fieldwork response rate here.

[30] Relative to the net PIAAC 2012 interviews, the response rate is 69 %.

[31] 2 473/4 214

[32] In our case, it is even possible that the head of household did not approve of the anchor person's participation in PIAAC.

5 Data management, Data Products, and Documentation

Various data processing activities took place at different stages. This included data cleaning, coding of open responses, and general data management activities. In addition, the competency measures from PIAAC Germany 2012 were re-scaled, yielding new plausible values to be used in the context of analyses with PIAAC-L data. A weighting strategy was determined for PIAAC-L, and subsequent weights produced. Finally, a number of data confidentiality edits were undertaken prior to the release of the data to the scientific community.

5.1 Data Cleaning

Basic data cleaning processes were carried out by both by the survey organisation and the PIAAC-L project partners, in various iterations.

First, the integrity and uniqueness of the sample (gross and net) and the identification numbers were checked and fixed, technical duplicates eliminated, and other such issues taken care of. This included a longitudinal check of key sociodemographic information with the corresponding PIAAC Germany 2012 data. The information about the other household members and adult household members who had participated in the PIAAC-L data collection was examined and it was ensured that the households were correctly identified and all allocations of the identification numbers were correct and consistent.

Second, the data processes usually run for the SOEP survey data on the net data were – as far as possible – reproduced for the PIAAC-L 2014 data (since essentially the same instruments were administered). However, certain adaptations to the PIAAC-L design were required. The first set of processes specific to the SOEP questionnaire were replicated by the SOEP team at the survey organisation (and not the PIAAC-L project director team, as for the other data management tasks) and primarily consisted in: (1) checking the functioning of the filters in the questionnaire, (2) identifying out of range responses, and (3) plausibility checks. Data errors were removed from the data at this stage, although no further contact was made with the respondents to obtain the correct responses. In addition, open entry responses were edited. Some of the open responses (i.e. country and nationality information) were also coded at this stage.

The second set of SOEP processes were carried out by the PIAAC-L project partners at the DIW, who had the main responsibility for producing the data sets for PIAAC-L 2014. The DIW group carried out additional data checks, generated derived variables, and generally prepared the data for publication. This included labeling the data in English and applying a missing scheme that was developed specifically for PIAAC-L.

Over its three waves of data collection, PIAAC-L brings together instruments from PIAAC, SOEP, and NEPS. These three surveys each have their own missing scheme tailored to each survey's specificities. For PIAAC-L a new missing scheme was developed to cater for the different data from the different surveys, since elements of each of the surveys, SOEP, PIAAC, and NEPS, are included at some time over the three waves of PIAAC-L. To do this, the missing schemes from PIAAC, SOEP, and NEPS were reviewed and a set of codes tailored to the PIAAC-L data was defined. The PIAAC-L missing scheme applied to the PIAAC-L 2014 data can be found in the PIAAC-L 2014 Preliminary Notes to the User (ZA5989_NotesToTheUser.pdf).[33] This missing scheme provides a brief description of each code

[33] Version accessible on 1.12.2016: https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5989&tdb=e&tdoi=10.4232/1.12576

(ranging from -1 to -9), and indicates the equivalent missing code in each of the sister surveys PIAAC, NEPS, and SOEP.

5.2 Coding

Open responses pertaining to occupation, industry, and certain education questions were coded by the coding department at the Research Data Center at the LIfBi. The coding of occupation and industry is described below. Coders were trained and their work was continuously monitored. In addition, the Research Data Center LIfBi coded vocational study programmes and degrees at university level as in the SOEP according to two German classifications FIELDS and DEGREE (cp. SOEP documentation, Goebel, 2014). Open language and country responses were coded by the survey organisation into the same country and language classifications used in the SOEP.[34]

Coding of Occupation and Industry

With respect to the information collected on occupation, it is important to consider the different approaches used in PIAAC 2012 and the SOEP survey. PIAAC aimed at obtaining occupation information coded into the *International Classification of Occupations* (ISCO-08; International Labour Office, 2012). The questions in the PIAAC background questionnaire were especially designed to elicit the range and amount of detail of responses to allow for coding into this international coding scheme. The person questionnaire for PIAAC-L 2014, however, was adopted from the SOEP core (see Section 3.3), and the questions on occupation (as well as some education questions) were focused on obtaining the kind of information required to code into the German National Classification of Occupations (Klassifikation der Berufe: KldB 2010; Bundesagentur für Arbeit, 2011). Thus, the respondent answers given to the occupation questions in PIAAC-L 2014 did not necessarily have the granularity required for reliable ISCO-08 coding.[35] Since this first wave of PIAAC-L data collection aimed at producing information based on SOEP standards, this was intended. However, for users of PIAAC 2012 data, an effort was made to generate additional ISCO-08 codes which would allow a general type of link with the PIAAC 2012 data. After coding into KldB 2010, a crosswalk to ISCO-08 was used covering approximately 90% of codes (cp. Bundesagentur für Arbeit, 2011). The remaining 10% of the codes could not be converted directly from KldB 2010 into ISCO-08 and were thus coded straight into ISCO-08.[36] Due to the fact that (a) different questions were administered, (b) a crosswalk was used, and (c) occupation coding was carried out by a different institution than in PIAAC, there are some limitations to the comparability of the PIAAC 2012 ISCO-08 codes and the PIAAC-L wave 1 codes.

Industry information was coded into the *International Standard Industrial Classification of All Economic Activities* (ISIC, Rev. 4; United Nations Statistics Division, 2013) as in PIAAC 2012. NACE codes were subsequently generated from these.

5.3 Re-Scaling of Plausible Values for PIAAC-L

The PIAAC 2012 assessment data was processed by the international PIAAC Consortium. They established separate proficiency scales for literacy, numeracy, and problem solving in technology-rich environments. This was achieved using Item Response Theory (IRT) scaling procedures and latent

[34] Some of these variables were coarsened according to a scheme developed for PIAAC Germany 2012 as a part of the data confidentiality edits described in Section 5.5.

[35] It should be noted that the underlying theoretical concepts of the KldB and ISCO are different.

[36] Coders were specifically trained for the direct coding into ISCO-08 and worked on many trial codings prior to their coding assignment. The trial codings were discussed and evaluated individually during training.

regression models yielding 10 plausible values for each assessment domain (for a detailed description of the PIAAC scaling procedures see Yamamoto, Khorramdel, & Von Davier, 2013a, 2013b, 2013c).

The first wave of data collection for PIAAC-L does not have a cognitive assessment component (a direct assessment is, however, a part of the second wave of data collection in 2015). However, by administering the elaborate SOEP-based questionnaires, the amount of background information on the anchor persons has been significantly increased. In order to be able to jointly analyse the newly collected information together with the competency measures and the background information from PIAAC 2012, it is necessary to extend the background model and produce a new set of plausible values. Re-scaling was carried out by the LIfBi: A new background model containing all the information included in PIAAC as well as the new PIAAC-L information was established, and plausible values were generated. The plausible values were released to enable analyses using competency measures from PIAAC 2012 and background questionnaires variables from both PIAAC-L 2014 and PIAAC 2012.[37] Some information on this re-scaling process for the first wave of PIAAC-L can be found in the Notes to the User mentioned in Section 5.1.

5.4 Weighting

Weighting for PIAAC-L combines the PIAAC and the SOEP weighting approaches. However, in comparison to the SOEP, for PIAAC-L weighting factors were only calculated for anchor persons, taking into account that sampling in PIAAC 2012 referred to a sample of individuals rather than a sample of households and their members. Weighting of the PIAAC-L 2014 data was carried out by the DIW and consisted in (1) modeling nonresponse in four steps, (2) generating nonresponse weights, (3) trimming, and (4) calibration. Calibration basically followed the PIAAC approach, but utilized a combination of raking and poststratification.[38] The replication approach followed in PIAAC, which accounted for the complex sampling design and thus was implemented to accurately calculate standard errors, was discontinued. For a detailed description of the weighting procedure in wave 1 see Bartsch and Poschmann (in press).

The weighting file[39] contains two final weighting factors for anchor persons: (1) the inverse staying probabilities, and (2) the poststratification weighting factors.

5.5 Data Confidentiality

The PIAAC-L Consortium is committed to adhering to survey ethics and undertaking any appropriate and effective measures required by data privacy concerns. These issues were already addressed by the international PIAAC project, and GESIS developed an elaborate data confidentiality strategy for the German PIAAC 2012 Scientific Use File.[40] The PIAAC-L strategy is generally based on this latter data confidentiality strategy for the national Scientific Use File. However, a number of additional aspects are taken into account: (1) Some of the specifics of the PIAAC-L design inevitably require some adaptations. For example, the extension to the household level excluded data privacy edits to household size (as was undertaken for PIAAC, where household size was top-coded). However, any deviations from the PIAAC strategy require a sound scientific justification. (2) Given the longitudinal

[37] These new plausible values are included in the person data set for PIAAC-L 2014 (ZA5989_Persons_14, see Section 5.6).

[38] PIAAC used poststratification only.

[39] ZA5989_Weights_14, see Section 5.6.

[40] A much more conservative strategy was also developed and implemented for the German PIAAC 2012 Public Use File.

approach of PIAAC-L, an impressive and extensive amount of data about the PIAAC Germany 2012 respondents is being amassed. This is considered during the decision-making process. (3) PIAAC-L is a cooperative undertaking and as such, all three institutes involved need to approve the data privacy measures. Even if only one party has a serious concern about releasing a certain variable or variables, this precludes the disclosure of this information. Thus, the final strategy may be more restrictive than if only one single institute was involved.

In developing and evaluating the necessary data confidentiality measures, a balance was sought between minimizing disclosure risk and minimizing information loss. Thus, the risk entailed by releasing more detailed information was weighed with the data utility and the scientific benefits. In the final consensual set of measures, a number of the PIAAC Germany 2012 data confidentiality edits were continued, such as the suppression of month of birth, suppression of the date of the interview, or the coarsening of country of birth and nationality. Some measures were tightened, for example PIAAC-L limited the amount of regional information even further than PIAAC Germany 2012 and released only information at the level of the Federal State.[41] Some measures were relaxed compared to PIAAC Germany 2012. For example, as mentioned above, household size was top-coded in PIAAC Germany 2012 Scientific Use File (at 10+), but due to the household approach followed by PIAAC-L, this restriction had to be discontinued.[42]

The data user contract represents another element towards ensuring data confidentiality. Only individuals who have personally signed the contract can obtain and use the data. The PIAAC-L data can only be used for previously specified research objectives, data usage length is restricted, and deletion of all data is mandatory at the end of usage. Data merging is generally forbidden with the exception of the PIAAC 2012 data.

Approval was required and obtained from all data protection officers and consultants as well as the principle investigators at each institute. The data confidentiality edits were implemented by GESIS and checked by the other project partners.

5.6 Data Products

One of the firm objectives of the PIAAC-L project is to provide academic researchers with an elaborate and high-quality data base with which to address a wide variety of research questions both in connection with the large-scale assessment data from PIAAC 2012, but also beyond. The data is therefore made available for scientific use by the Research Data Centre PIAAC (RDC PIAAC/FDZ PIAAC)[43] via the GESIS Data Catalogue. The RDC PIAAC offers various PIAAC and PIAAC-related data sets, provides support to users and training workshops for PIAAC data users.

In order to obtain access to the data, it is necessary to register and sign the specific PIAAC-L Data Use Agreement. The latter encompasses all three waves of data collection and users are automatically notified of new PIAAC-L data releases. The PIAAC-L data release includes the delivery of the PIAAC Germany 2012 Scientific Use File.[44] The PIAAC-L user contract explicitly allows the PIAAC-L data files

[41] In the PIAAC Scientific Use File, regional information was provided at an aggregated level.

[42] However, for households with 10 or more persons, variables containing ethnic information were scrutinized and anonymized at an individual level as required.

[43] http://www.gesis.org/en/piaac/rdc/

[44] Under a separate study number and doi.

to be merged with the PIAAC Germany 2012 Scientific Use File and thus enables longitudinal analyses.[45]

The released PIAAC-L scientific use database for the first wave of data collection in 2014 consists of five separate data sets (GESIS Data Archive, Data File ZA5989, Version accessible on 1.12.2016: 1.1.0, *doi:10.4232/1.12576*):

- ZA5989_Persons_14
  - → Units: All PIAAC-L 2014 respondents (anchor persons and household members 18+ with participation in data collection 2014)
  - → Content: Data from person questionnaire, including derived variables
- ZA5989_Household_14
  - → Units: All PIAAC-L 2014 households (households are defined by participating anchor persons)
  - → Content: Data from household questionnaire, including derived variables
- ZA5989_Calendar
  - → Units: All PIAAC-L 2014 respondents (anchor persons and household members 18+ with participation in data collection 2014)
  - → Content: Data from biographical calendar, spell data
- ZA5989_Registry
  - → Units: All persons ever registered in PIAAC-L
  - → Content: Basic information on participation in the different waves of data collection
- ZA5989_Weights_14
  - → Units: Anchor persons 2014
  - → Content: Weighting factors

Various pieces of documentation are also available: A documentation of the questionnaire (in German, as administered in the field, but with English labels), an extensive codebook for each data set (in English), and notes to the users.[46] In addition, there are technical reports as well as the survey organisation's fieldwork report (in German only).

[45] The Data User Contract for the PIAAC Germany 2012 Scientific Use File prohibits merging with other personal data.

[46] The Notes to the Users is a *growing documentation* that is elaborated on by the PIAAC-L group with time.

6 Conclusion

The first wave of PIAAC-L has successfully achieved the transition from the international cross-sectional survey PIAAC to the national longitudinal survey PIAAC-L. Obtaining respondent participation was especially crucial in this first data collection wave, since anchor persons who were not successfully recruited were lost for the entire project (as were their households). The efforts invested in fieldwork were well placed, as reflected in the 72% response rate. This result can be regarded as very successful, especially considering (1) that PIAAC was originally set up as a cross-sectional survey, (2) the amount of time elapsed between PIAAC Germany 2012 and the first wave of data collection for PIAAC-L, as well as (3) the extended perspective of PIAAC-L (both in terms of the longitudinal character as well as the extension to other household members). While there are many challenges in reconciling the different characteristics in PIAAC and – in the case of this first wave of data collection – the SOEP, this has been a very constructive and enlightening process. As a result, the PIAAC-L 2014 data constitute an empirically sound body of evidence which significantly extends and enriches the PIAAC Germany 2012 data. The comprehensive information on respondents and households in the data sets for this first wave of PIAAC-L has been released and made accessible to the scientific community.

Annex: Questions Excluded From Original SOEP Instruments in the PIAAC-L Wave 1 Person Questionnaire

The construction of the PIAAC-L 2014 person questionnaire was based on the SOEP *Biography Questionnaire* and the SOEP *Individual Questionnaire*.

The following questions were excluded from original SOEP *Biography Questionnaire* (TNS Infratest Sozialforschung, 2014b):

- Questions on immigration[47]
- Questions on nationality[48] were omitted because they were already included in the *Individual Questionnaire.*

The following questions were excluded from the SOEP *Individual Questionnaire* (TNS Infratest Sozialforschung, 2014c):

- Questions about working hours and working time regulations[49]
- A sub-set of questions pertaining to secondary employment[50]
- Follow-up questions regarding last job[51,52]
- Employment calendar for the year 2013[53,54]
- Questions about health insurance[55]
- Left-right scale[56]
- Risk propensity: Willingness to take risks in different areas[57]
- Attachment to local area/region, transnationalization[58]
- Other sets of questions which overlapped with the *Biography Questionnaire* were omitted (e.g. educational qualifications obtained in 2013)[59]

[47] Questions 6-13
[48] Questions 14-19
[49] Questions 43-46
[50] Questions 61-63
[51] Were already included from the *Biography Questionnaire.*
[52] Questions 80-83
[53] Were already included (on a monthly basis) from the *Biography Questionnaire.*
[54] Question 84
[55] Questions 105, 107-117
[56] Question 122
[57] Question 125
[58] Questions 140-148
[59] Questions 66-69

References

Bartsch, S., & Poschmann, K. (in press). Weighting in PIAAC-L. Köln: GESIS Papers.

Bundesagentur für Arbeit. (2011). *Klassifikation der Berufe 2010 – Band 1: Systematischer und alphabetischer Teil mit Erläuterungen*. Nürnberg: Bundesagentur für Arbeit. Retrieved from https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB-2010/Generische-Publikationen/KldB2010-Printversion-Band1.pdf

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087-1101. doi:10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short Grit Scale (Grit–S). *Journal of Personality Assessment, 91*(2), 166 – 174. doi:10.1080/00223890802634290

Glemser, A., Huber, S., & Bohlender, A. (TNS Infratest Sozialforschung) (2016). *SOEP 2014 – Methodenbericht zum Befragungsjahr 2014 (Welle 31) des Sozio-oekonomischen Panels*. SOEP Survey Papers 299: Series B. Berlin: DIW/SOEP. Retrieved from http://www.diw.de/sixcms/detail.php?id=diw_01.c.527824.de

Goebel, J. (2014). *SOEP Documentation - Codebook v29: Generated variables – personal level*. Retrieved from https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.437095.de/codebook_v29_pgen_klein.pdf

International Labour Office. (2012). *International Standard Classification of Occupations ISCO-08: Structure, group definitions and correspondence tables* (Vol. 1). Geneva: International Labour Organization.

Mohadjer, L., Krenzke, T., & Van de Kerckhove, W. (2013a). Sampling design. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 14, pp. 1-31). Paris: OECD. Retrieved from http://www.oecd.org/skills/piaac/publications.htm

Mohadjer, L., Krenzke, T., & Van de Kerckhove, W. (2013b). Indicators of the quality of the sample data. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 16, pp. 1-30). Paris: OECD. Retrieved from http://www.oecd.org/skills/piaac/publications.htm

OECD. (2010). PIAAC technical standards and guidelines. Retrieved March 15, 2014 from http://www.oecd.org/site/piaac/surveyofadultskills.htm

OECD. (2011). *PIAAC conceptual framework of the background questionnaire main survey*. Paris: OECD.

OECD. (2013). *OECD skills outlook 2013: First results from the Survey of Adult Skills*. Paris: OECD.

Rammstedt, B. (2013). *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich – Ergebnisse von PIAAC 2012*. Münster: Waxmann.

Rammstedt, B., Ackermann, D., Helmschrott, S., Klaukien, A., Maehler, D. B., Martin, S., Massing, N., & Zabal, A. (2013a). *PIAAC 2012. Die wichtigsten Ergebnisse im Überblick.* Münster: Waxmann.

Rammstedt, B. , Ackermann, D., Helmschrott, S., Klaukien, A., Maehler, D. B., Martin, S., Massing, N., & Zabal, A. (2013b). *PIAAC 2012. Overview of the main results.* Münster: Waxmann.

Rammstedt, B., Martin, S., Zabal, A., Carstensen, C., & Schupp, J. (in press). The PIAAC Longitudinal Study in Germany - rationale and design. *Large-scale Assessments in Education.*

Steinacker, G., Schmidt, S., Wolfert, S., & Schneekloth, U. (2016). *Durchführung der 1. Erhebungswelle von PIAAC-L (Kooperative längsschnittliche Weiterverfolgung der PIAAC-Studie in Deutschland).* Köln: GESIS Papers 2016|05.

TNS Infratest Sozialforschung. (2014a). *SOEP 2014 – Erhebungsinstrumente 2014 (Welle 31) des Sozio-oekonomischen Panels: Haushaltsfragebogen, Altstichproben.* SOEP Survey Papers 236: Series A. Berlin: DIW/SOEP Retrieved from https://www.diw.de/de/diw_01.c.389326.de/publikationen_veranstaltungen/publikationen/soep_survey_papers/soep_survey_papers.html?gsid=diw_01.c.389326.de&tid=389326&ty=2014&tskip=20

TNS Infratest Sozialforschung. (2014b). *SOEP 2014 – Erhebungsinstrumente 2014 (Welle 31) des Sozio-oekonomischen Panels: Lebenslauffragebogen, Altstichproben.* SOEP Survey Papers 237: Series A. Berlin: DIW/SOEP. Retrieved from https://www.diw.de/de/diw_01.c.389326.de/publikationen_veranstaltungen/publikationen/soep_survey_papers/soep_survey_papers.html?gsid=diw_01.c.389326.de&tid=389326&ty=2014&tskip=20

TNS Infratest Sozialforschung. (2014c). *SOEP 2014 – Erhebungsinstrumente 2014 (Welle 31) des Sozio-oekonomischen Panels: Personenfragebogen, Altstichproben.* SOEP Survey Papers 235: Series A. Berlin: DIW/SOEP. Retrieved from https://www.diw.de/de/diw_01.c.389326.de/publikationen_veranstaltungen/publikationen/soep_survey_papers/soep_survey_papers.html?gsid=diw_01.c.389326.de&tid=389326&ty=2014&tskip=20

United Nations Statistics Division. (2013). Detailed structure and explanatory notes ISIC Rev. 4. Retrieved July 12, 2013 from http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27

Yamamoto, K., Khorramdel, L., & Von Davier, M. (2013c). Scaling PIAAC cognitive data. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 17, pp. 1-33). Paris: OECD. Retrieved from http://www.oecd.org/skills/piaac/publications.htm

Yamamoto, K., Khorramdel, L., & Von Davier, M. (2013b). Scaling outcomes. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 18, pp. 1-28). Paris: OECD. Retrieved from http://www.oecd.org/skills/piaac/publications.htm

Yamamoto, K., Khorramdel, L., & Von Davier, M. (2013a). Proficiency scale construction. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 19, pp. 1-22). Paris: OECD. Retrieved from http://www.oecd.org/skills/piaac/publications.htm

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