

The Research Data Center PIAAC at GESIS

Perry, Anja; Rammstedt, Beatrice

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Perry, A., & Rammstedt, B. (2016). The Research Data Center PIAAC at GESIS. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics*, 236(5), 581-593. <https://doi.org/10.1515/jbnst-2015-1024>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/3.0>

Anja Perry and Beatrice Rammstedt*

The Research Data Center PIAAC at GESIS

DOI 10.1515/jbnst-2015-1024

1 Introduction

With the Programme for the International Assessment of Adult Competencies (PIAAC) researchers can shed light on how competencies are acquired, how its use helps us maintain and further develop skills, and whether adults are prepared for the challenges of modern knowledge societies (OECD 2013a). The Organisation for Economic Co-operation and Development (OECD) initiated PIAAC in more than 30 countries to assess competencies of the adult population. Similar to the Programme for International Student Assessment (PISA), PIAAC is planned to be repeated in regular intervals. Therefore, the next cycle of PIAAC is planned for 2022.

The OECD published the PIAAC international public use file of the first cycle of PIAAC (OECD 2015) in 2013. Due to German confidentiality rules, GESIS published a scientific use file (Rammstedt et al. 2015) that includes information that could not be released in the public use file. Further national data and para data for PIAAC can shed light on further research questions as well as methodological aspects of PIAAC. This data is currently and will be made available in the Research Data Center PIAAC (RDC PIAAC) at GESIS. In addition to this, various add-on studies were and are currently being conducted in Germany, such as Competencies in Later Life (CiLL) and the longitudinal study PIAAC-L.

However, the PIAAC data presents challenges due to imputed competency scores (plausible values) and country-specific complex sample techniques. The RDC PIAAC provides information on analytic methods and the available analysis tools. It also offers workshops to familiarize users with the data and to teach them how to analyze the PIAAC data.

Given the brevity of the PIAAC data release, an impressive number of research papers were published that use PIAAC data. Research with PIAAC focuses, for example, on the returns to skills (e. g., Hanushek et al. 2015), skill

***Corresponding author: Beatrice Rammstedt**, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany, E-mail: beatrice.rammstedt@gesis.org

Anja Perry, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany, E-mail: anja.perry@gesis.org

and wage inequality (Paccagnella 2015), skill mismatch (Allen et al. 2013a; Perry et al. 2014), non-monetary outcomes, such as trust (Borgonovi/Burns 2015), and also methodological aspects, such as incentives in large-scale assessments (Martin et al. 2014).

This paper aims to present central aspects of PIAAC, analytical procedures for the competence measures and the complex sample design, as well as data, information and services provided through the RDC PIAAC at GESIS.

2 Assessing adult competencies around the world – The PIAAC sample and country coverage

PIAAC aims to assess adult competencies. To do so, a sample of 16 to 65 year old individuals in private households, irrespective of their nationality and legal status, was drawn in each participating country. Depending on what is available in each country, either a household survey or a registry-based survey was drawn. The target response rate in each country was 70 %, the minimum response rate was 50 % with at least 5,000 respondents. In Germany, a two-stage stratified and clustered random sample was drawn based on municipality registries. The final German sample comprised 5,465 adults thus representing a response rate of 55 % (Mohadjer et al. 2013a; Zabal et al. 2014).

Twenty-four countries participated in the first round of PIAAC. The data for this round was published in 2013. In order to extend the scope of PIAAC, the OECD initiated further rounds to include additional countries. Table 1 gives an overview of participating countries in each round. Data from round 2 will be published in 2016.

Table 1: Countries participating in PIAAC.

| | |
|--------------------------------|---|
| PIAAC Round I: | Australia, Austria, Canada, Cyprus, Czech Republic, Denmark, England/N. Ireland (UK), Estonia, Finland, Flanders (Belgium), France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Spain, Sweden, United States |
| PIAAC Round II: | Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia, Turkey |
| PIAAC Round III (preliminary): | Kazakhstan, Mexico, Peru, Hungary, Ecuador |

3 Background information and proficiency measured with PIAAC

3.1 The PIAAC background questionnaire

Prior to the cognitive assessment in PIAAC, respondents were administered a detailed background interview. This was a computer-assisted personal interview (CAPI) and took 30 to 40 min (Kirsch/Yamamoto 2013).

In addition to the assessment of socio-demographic and socio-economic information, the background questionnaire was intended to shed light on various research questions (OECD 2011). These research questions concerned the acquisition and loss of skills, skill use, skill inequality, human capital theory, signaling theory, and skills outcomes (Allen et al. 2013b).

The CAPI mode allows for complex routings during the interview, which decreases the burden for the respondent because questions can be easily skipped if they do not apply. Figure 1 shows a simplified routing scheme by background questionnaire sections. More complex routing rules apply within some sections of the questionnaire.

3.2 The cognitive assessment in PIAAC

The central part of the PIAAC study was the cognitive assessment, which followed the background interview. At this stage of the interview process, the interviewer handed the laptop to the respondent for him or her to conduct the self-directed assessment. In case the respondent was routed to the paper branch, he or she received paper booklets with the tasks to be solved. The interviewer was not allowed to help the respondent in any way. Only an electronic calculator and ruler were provided.

Three independent expert groups developed the theoretical frameworks of the proficiency domains tested in PIAAC: literacy (Jones et al. 2009), numeracy (Gal et al. 2009), and problem-solving in technology-rich environments (Rouet et al. 2009). Findings and experience from the previous large-scale assessments International Adult Literacy Survey (IALS), the Adult Literacy and Life Skills Survey (ALL) as well as the Programme for the International Student Assessment (PISA) were taken into account during framework development.

By default, the assessment of these domains was administered on the computer. However, when respondents lacked computer experience, failed a simple ICT test, or refused to take the test on the computer, they could also do

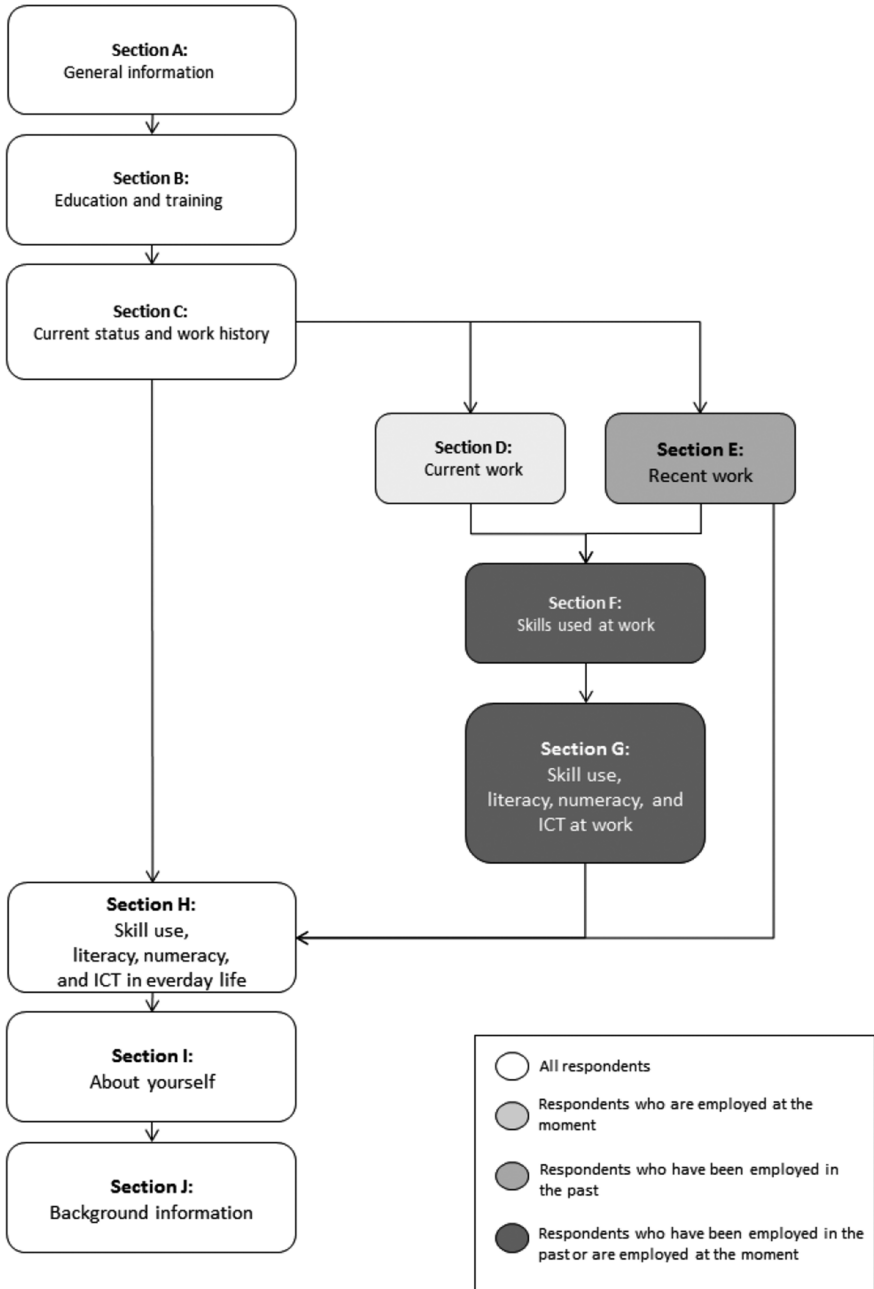


Figure 1: General routing of the background questionnaire (Zabal et al., 2014: 22).

a paper-based assessment. In this case, only literacy and numeracy were assessed. In the paper branch, respondents randomly received either literacy or numeracy tasks. Problem-solving in technology-rich environments are, by definition, computer-based, so that this domain is not covered by the paper branch. Assuming that mostly respondents with low skills are selected into the paper branch, a short assessment of the respondents' reading components, a test of basic reading skills (Sabatini/Bruce 2009), was added to the paper-based assessment.

4 Item response theory and replicate approach – The challenges of the data analysis with PIAAC

Respondents in PIAAC were administered only a subset of the pool of available assessment tasks and only one or two out of the three competency domains. This reduction was necessary to decrease the burden for the respondents and increase the motivation to participate in PIAAC.

In order to draw meaningful conclusions on the populations' proficiency, the missing values were imputed. Item Response Theory and a latent regression model (Mislevy 1984; Mislevy 1985) were then applied to estimate consistent proficiency measures on the population level. Doing so, all information from the background questionnaire was combined with the information from the respondent's cognitive assessment to derive a posterior distribution that represents the individual's proficiency (Yamamoto et al. 2014).

To account for the uncertainty added by multiple imputation, 10 so called plausible values are drawn from the posterior distribution instead of one single proficiency score. These 10 plausible values have to be accounted for using Rubin's Rule (Rubin 1987). According to this rule, each estimate is calculated with each plausible value individually. The average of the resulting estimates produces an estimate and confidence interval that accounts for the uncertainty due to missing data. Rutkowski et al. (2010) provide a detailed description on how to handle multiple imputations in large-scale assessment data.

To account for the complex sample design applied in PIAAC (e. g., stratification and clustering, Mohadjer et al. 2013b), subsamples (replicate samples) were drawn from the total sample and weights were assigned to each subsample. The number of subsamples and replicate weights differ across countries, yielding up to 80 in most countries. These replicate weights must be taken into account

in order to draw meaningful conclusions for the population, using the PIAAC data. Mohadjer et al. (2013b) provide detailed information about the replicate approach used in PIAAC.

In sum, to account for the complex sample design one needs to calculate each statistic 80 times (in some countries fewer times) while taking the different replicate weights into consideration. Together with one additional calculation using the final sample weight, this results in 81 calculations when all weights are taken into account and 810 calculations when all plausible values (proficiency) are taken into account also. Further information on the use of plausible values and replicate weights when analyzing the PIAAC data can be found in the User Guide for the German PIAAC Scientific Use File (Perry/Helmschrott 2014).

5 Data, information, and service at the research data center PIAAC

Along with the PIAAC results and two extensive reports (OECD 2013a, 2013b) the OECD published the international public use file for 22 countries (OECD 2015).¹ The public use file is an extensive data base that offers cross-country analyses on a detailed level. As it is freely available for everyone, it is also a very extensive source for, for example, teaching environments. However, because of different data protection rules in the participating countries, not all information collected in PIAAC is available in the public use file. GESIS therefore published a scientific use file for Germany to allow detailed analyses for these countries.

With the RDC PIAAC we aim to provide extensive German data in addition to the public use file and further international data collected during the PIAAC project. Figure 2 shows current and planned data available at the RDC PIAAC. The data sets in white indicate data currently available at the RDC PIAAC in March 2016. The data sets displayed in grey are planned to be published in the future. This includes extensive survey and para data useful for methodological research. Further information on the available data at the RDC PIAAC can be found on the German PIAAC website www.gesis.org/piaac.

¹ Although Australia and Cyprus participated in PIAAC 2012, their data are not included in the public use file. The Australian data is available at the Australian Bureau of Statistics. To access the data, please write to microdata.access@abs.gov.au. The public use file for Cyprus is available in the RDC PIAAC at GESIS (see below; Michaelidou-Evripidou et al. (2014).

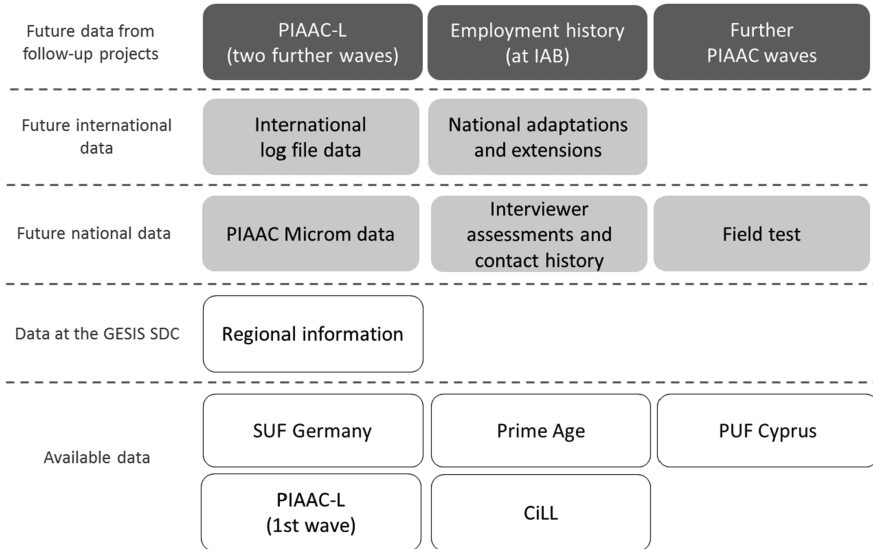


Figure 2: Current and planned data supply by the RDC PIAAC.

Note: SDC = Secure Data Center at GESIS, IAB = Institute for Employment Research.

5.1 Data currently available at the RDC PIAAC

The following data is currently available at the RDC PIAAC:

- 1) The *German PIAAC Scientific Use File (SUF)* was published in March 2014 and is available for researchers after signing a data user agreement (see below). The German PIAAC SUF can be easily merged with the public use files of other countries available at the OECD homepage (OECD 2015). In the PIAAC SUF User Guide we provide a merge syntax (Perry/Helmschrott 2014).
- 2) The *PUF for Cyprus* is freely available at the RDC PIAAC. It is not provided by the OECD.
- 3) Data on *prime age workers* (26 and 55 years old) in Germany is available in the RDC PIAAC as SUF. This data results from a German PIAAC supplement study by the Research Center Berlin (WZB), which investigates labor market success of 26 to 55 year old adults who are low educated. The data file includes all PIAAC respondents between 26 and 55 years of age, including an oversample of 560 adults of the same age group in East Germany.
- 4) The supplement study „*Competencies in later life*“ (CiLL),² conducted by the German Institute for Adult Education – Leibniz Centre for Lifelong

² Further information: <http://www.die-bonn.de/cill/default.aspx>.

Learning (DIE) and Ludwig-Maximilians-University Munich (LMU), surveyed adults from 66 to 80 years old and assessed their competencies. The CiLL SUF is available at the RDC PIAAC.

- 5) In Germany, the PIAAC sample was transferred into the panel PIAAC-L. This is a joint project of GESIS, the Leibniz Institute for Educational Trajectories (LifBi) and the German Institute for Economic Research (DIW), financed by the Federal Ministry of Education and Research. PIAAC-L offers a wide scope of analyses regarding the longitudinal development of skills, additional background information and relationships between partners and household members. Data from the first PIAAC-L wave is currently available at the RDC PIAAC as a SUF. In the following years, the data of the two remaining waves of PIAAC-L will be published for scientific use.

5.2 Regional information accessible at the GESIS secure-data-center

PIAAC data on a very detailed regional level can provide further insights into cognitive skills, such as returns to ICT skills (Falck et al. 2016). For data confidentiality reasons, however, this information is available in the GESIS Secure-Data-Center (SDC) in Cologne.³ Researchers can access this data on-site at GESIS and analyze the data in a secure environment. Only aggregated results are then provided to the researcher. *Regional information*, such as municipality size and type of agglomeration (BIK) is available for the place of residence of the PIAAC respondents. This type of data can be used for spatial analyses as well as for non-response analyses.

5.3 Additional PIAAC data to be published

- 1) Data from Microm Consumer Marketing⁴ was used to tailor notifications to certain target groups, such as the low-educated, and for non-response-bias analyses (Helmschrott/Martin 2014; Zabal et al. 2014). Microm provides geo-marketing data at the street level and was added to the PIAAC data based on respondents' addresses. This data allows for further analyses, especially

³ Further information: <http://www.gesis.org/unser-angebot/daten-analysieren/datenservice/secure-data-center-sdc/>.

⁴ Further information: <http://www.microm-online.de/>.

regarding respondents' neighborhoods, environments, as well as sampling analyses.

- 2) *Interviewer assessments and contact documentations* can be used for analyses of contact histories and the respondents' living environments or can be used as additional (control) variables when analyzing PIAAC.
- 3) *Log File Data* from the computer-based competency assessment in PIAAC offer important insights in strategies used when solving the assessment tasks. In a joint project of GESIS and the German Institute for International Educational Research (DIPF) and financed by the OECD, a tool is developed to extract data from the log files into a flat data file. Along with the log file data, the cognitive assessment tasks of PIAAC will be documented and made available for the users.
- 4) Each country was able to add *country-specific adaptations and extensions*. This country specific data is often not available to researchers, as they were not included in the international PUF. GESIS plans to combine all country specific data in order to make it available to the users through the RDC.
- 5) An extensive *field test* for PIAAC took place in 2010 and its data will be made available for researchers. In addition to further background variables, the selection into the paper based branch was random. This allows research on mode effects and improved analyses of the reading components.
- 6) Employment biography data of the PIAAC respondents will be linked to the PIAAC data. The linking of the PIAAC data with the employment biography data is executed by the IAB and will be accessible at any on-site use location of the IAB across Germany and the USA.⁵ The RDC PIAAC can be contacted for consultation and for establishing contact with the IAB.

Furthermore, PIAAC will be repeated in a 10-year rhythm. We plan to also provide data from future PIAAC waves through the RDC PIAAC.

5.4 Documentation

Besides the large amount of data that will be made available for the scientific community, the RDC offers a comprehensive documentation of the available data, such as codebooks, questionnaires, and technical reports (Zabal et al. 2014). In addition, the variables of each PIAAC survey will be entered into ZACAT, an online study catalogue. This social science data portal allows users

⁵ For IAB on-site access locations visit http://fdz.iab.de/en/FDZ_Data_Access/FDZ_On-Site_Use/Standorte.aspx.

to search for, browse, analyze, and download social science survey data. It also provides documentation of full question and answer texts, on the variable level. In addition, the RDC PIAAC provides information on the correct usage of plausible values and replicate weights when analyzing PIAAC and on the available analysis tools for different analytic problems (Perry/Helmschrott 2014).

5.5 Data training

Data workshops will be offered to familiarize current and potential users with the PIAAC data and teach them how to competently analyze the data. A first workshop in October 2015, financed by the Federal Ministry of Education and Research, focused on data access, the handling of complex sampling design and multiple imputation, and the analyses of PIAAC in a multilevel design. Further workshops with different methodological foci will take place in the future and will be announced on the PIAAC website at GESIS (www.gesis.org/piaac).

5.6 Consultation

Users interested in the PIAAC data, having questions regarding data access or regarding data analyses with PIAAC can contact the RDC PIAAC via email (fdz-piaac@gesis.org).

6 Data access at the research data center PIAAC

Users who wish to have access to the PIAAC data need to register with the GESIS data archive at <https://dbk.gesis.org/register/index.asp?db=D>

They will then be able to access all RDC PIAAC data from there. For most data sets users need to sign a data user agreement in order to verify their status as a researcher. This user agreement must be sent to the number or address stated on the agreement via fax or mail. The data can then be accessed through a secure link that will be provided usually within one working day. The public use file for Cyprus is available for all users and no user agreement needs to be signed. This data can be downloaded directly after registration.

Currently, a fee of 20€ is charged whenever data is requested that is restricted to scientific users and requires a signed user agreement. This fee includes the processing of a total of five restricted data sets and can be used

for any other data provided by the GESIS data archive. Data freely available for all user groups, f.ex., the PIAAC Public Use File for Cyprus, is free of charge.

Users who wish to access data in the SDC can contact the GESIS SDC by email (sdc@gesis.org). The GESIS SDC will then schedule a time slot in order to facilitate data access at GESIS Cologne and the availability of user support by the RDC PIAAC. Users can bring their own material and data to the GESIS SDC. However, all incoming material will be checked beforehand to ensure that no re-identification of respondents is possible. After the users completed their analyses, outgoing data will be checked again for potential re-identification threats and then sent to the users.

7 Summary

PIAAC provides recent data on cognitive skills and allows investigating the role they play in various aspects, such as in education and on the labor market as well as regarding non-monetary outcomes. As presented, a number of data sets are available in the RDC PIAAC. Along with para data collected during the field process, the PIAAC scientific use file can open up a wide scope of research avenues in both methodological and topical research fields.

The application of item response theory and the replicate approach imply certain challenges when analyzing the PIAAC data. Therefore the RDC PIAAC provides extensive documentation on the PIAAC data and offers workshops in which data users learn how to competently analyze the data. The RDC PIAAC can also be contacted for questions regarding data analysis and data access.

References

- Allen, J., Levels, M., van der Velden, R. (2013a), Skill Mismatch and Skill Use in Developed Countries: Evidence from the PIAAC Study. ROA Research Memorandum, Maastricht, Research Centre for Education and the Labour Market (ROA).
- Allen, J., van der Velden, R., Helmschrott, S., Martin, S., Massing, N., Rammstedt, B., Zabal, A., Von Davier, M. (2013b), The Development of the PIAAC Background Questionnaires. in: OECD (ed.), Technical Report of the Survey of Adult Skills (PIAAC) – Pre-publication copy. Paris, OECD.
- Borgonovi, F., Burns, T. (2015), The Educational Roots Of Trust. OECD Education Working Papers. Paris, OECD Publishing.
- Falck, O., Heimisch, A., Wiederhold, S. (2016), Returns to ICT Skills. CESifo Working Paper 5720.

- Gal, I., Alatorre, S., Close, S., Evans, J., Johansen, L., Maguire, T., Manly, M., Tout, D. (2009), PIAAC Numeracy: A Conceptual Framework. OECD Education Working Paper No. 35. Paris, OECD Publishing.
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., Woessmann, L. (2015), Returns to Skills Around the World: Evidence from PIAAC. *European Economic Review* 73 (C): 103–130.
- Helmschrott, S., Martin, S. (2014), Nonresponse in PIAAC Germany. *Methods, Data, Analyses* 8 (2): 243–266.
- Jones, S., Gabrielsen, E., Hagston, J., Linnakylä, P., Megherbi, H., Sabatini, J., Tröster, M., Vidal-Abarca, E. (2009), PIAAC Literacy: A Conceptual Framework. OECD Education Working Paper No. 34. Paris, OECD Publishing.
- Kirsch, I., Yamamoto, K. (2013), PIAAC Assessment Design. In: OECD (ed.), *Technical Report of the Survey of Adult Skills (PIAAC) – Pre-publication copy*. Paris, OECD.
- Martin, S., Helmschrott, S., Rammstedt, B. (2014), The Use of Respondent Incentives in PIAAC: The Field Test Experiment in Germany. *Methods, Data, Analyses* 8 (2): 223–242.
- Michaelidou-Evripidou, A., Modestou, M., Karagiorgi, Y., Polydorou, A., Nicolaidou, M., Afantiti-Lamprinou, T., Kendeou, P., Tsouris, C., Loukaides, C. (2014), Programme for the International Assessment of Adult Competencies (PIAAC), Cyprus. GESIS Data Archive. Cologne, Germany. ZA5650 Data file Version 1.0.0, doi:10.4232/1.11906.
- Mislevy, R. J. (1984), Estimating Latent Distributions. *Psychometrika* 49 (3): 359–381.
- Mislevy, R. J. (1985), Estimation of Latent Group Effects. *Journal of the American Statistical Association* 80 (392): 993–997.
- Mohadjer, L., Krenzke, T., Van de Kerckhove, W. (2013a), Sampling Design. in: OECD (ed.), *Technical Report of the Survey of Adult Skills (PIAAC) – Pre-publication copy*. Paris, OECD.
- Mohadjer, L., Krenzke, T., Van de Kerckhove, W. (2013b), Survey Weighting and Variance Estimation. in: OECD (ed.), *Technical Report of the Survey of Adult Skills (PIAAC) – Pre-publication copy*. Paris, OECD.
- OECD (2011), PIAAC Conceptual Framework of the Background Questionnaire Main Survey. Paris, OECD.
- OECD (2013a), OECD Skills Outlook 2013: First Results From the Survey of Adult Skills. Paris, OECD Publishing.
- OECD (2013b), *The Survey of Adult Skills: Reader's Companion*. Paris, OECD Publishing.
- OECD (2015), Programme for the International Assessment of Adult Competencies (PIAAC), International Public Use File. Paris, France, OECD.
- Paccagnella, M. (2015), Skills and Wages Inequality: Evidence from PIAAC. OECD Education Working Papers. Paris, OECD Publishing.
- Perry, A., Helmschrott, S. (2014), User Guide for the German PIAAC Scientific Use File. GESIS. Available at: <https://dbk.gesis.org/dbksearch/download.asp?db=D&id=52078>.
- Perry, A., Wiederhold, S., Ackermann-Piek, D. (2014), How Can Skill Mismatch Be Measured? New Approaches with PIAAC. *Methods, Data, Analyses* 8 (2): 137–174.
- Rammstedt, B., Zabal, A., Martin, S., Perry, A., Helmschrott, S., Massing, N., Ackermann, D., Maehler, D. (2015), Programme for the International Assessment of Adult Competencies (PIAAC), Germany – Reduced Version. GESIS Data Archive. Cologne Germany. ZA5845 Data file Version 2.1.0, doi:10.4232/1.12385.
- Rouet, J. -F., Bétrancourt, M., Britt, M. A., Bromme, R., Graesser, A. C., Kulikowich, J. M., Leu, D. J., Ueno, N., van Oostendorp, H. (2009), PIAAC Problem Solving in Technology-Rich

- Environments: A Conceptual Framework. OECD Education Working Paper No. 36. Paris, OECD Publishing.
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*. New York, NY, John Wiley & Sons.
- Rutkowski, L., Gonzalez, E., Joncas, M., von Davier, M. (2010), *International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting*. *Educational Researcher* 39 (2): 142–151.
- Sabatini, J. P., Bruce, K. M. (2009), *PIAAC Reading Components: A Conceptual Framework*. OECD Education Working Paper No. 33. Paris, OECD Publishing.
- Yamamoto, K., Khorramdel, L., Von Davier, M. (2014), *Scaling PIAAC Cognitive Data*. Paris, OECD.
- Zabal, A., Martin, S., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I., Rammstedt, B. (2014), *PIAAC Germany 2012: Technical Report*. Münster, Waxmann.