

## Number Series Study (DIPF): Technical Report

Engelhardt, Lena; Goldhammer, Frank

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

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

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Engelhardt, L., & Goldhammer, F. (2018). *Number Series Study (DIPF): Technical Report*. (GESIS Papers, 2018/01). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.55737>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nc/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more information see: <https://creativecommons.org/licenses/by-nc/4.0>

gesis

Leibniz-Institut  
für Sozialwissenschaften

GESIS Papers

2018|01

**Number Series Study (DIPF)**

**Technical Report**

*Lena Engelhardt & Frank Goldhammer*



GESIS Papers 2018|01

## Number Series Study (DIPF)

### Technical Report

*Lena Engelhardt<sup>1</sup> & Frank Goldhammer<sup>1,2</sup>*

<sup>1</sup> German Institute for International Educational Research (DIPF),  
Frankfurt am Main, Germany

<sup>2</sup> Centre for International Student Assessment (ZIB), Germany

## **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](mailto:anouk.zabal@gesis.org)

**Correspondence:** Lena Engelhardt / Frank Goldhammer, DIPF Frankfurt, German Institute for International Educational Research, Schloßstr. 29, 60486 Frankfurt am Main, Phone: +49 69 24 708 754, Email: [engelhardtl@dipf.de](mailto:engelhardtl@dipf.de); [goldhammer@dipf.de](mailto:goldhammer@dipf.de)

**Acknowledgments:** We would like to thank Anouk Zabal and Silke Martin for their helpful comments on a first draft of this paper.

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

## Contents

---

1	Introduction .....	5
2	The PIAAC-L Sample for the Number Series Study.....	8
3	Instruments.....	9
3.1	Number Series .....	9
3.2	Strategy Questions.....	10
4	Fieldwork.....	12
5	Data Management, Data Products, and Documentation.....	13
6	Conclusion .....	15
	References.....	16



## 1 Introduction

---

The Number Series Study addresses the fundamental research question of what is measured by competence tests used in the Survey of Adult Skills (PIAAC; Programme for the International Assessment of Adult Competencies): More specifically, to what extent do the PIAAC Literacy and Numeracy measures capture the outcome of learning processes, and to what extent do they measure general cognitive abilities?

PIAAC competencies are understood as key information-processing competencies and assumed to be necessary to master various everyday situations, to be transferable, and to be learnable (OECD, 2016, p. 16). Relating PIAAC competence scores to other variables, for instance economic and social outcomes, may shed some light on how education can enable prosperity for individuals and groups. However, such important conclusions require that the PIAAC competence tests actually measure learning outcomes from educational settings in the first place and not general cognitive abilities depending on factors outside educational settings.

Thus, the intended interpretation of test scores obtained from a competence assessment needs to be supported both theoretically and empirically (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014). Validity evidence is needed for the claim that differences in proficiency scores from the PIAAC competence assessments represent differences in learning outcomes. One major approach to investigate what PIAAC proficiency scores actually indicate – merely general cognitive abilities or also learning outcomes – is to compare these proficiency scores with test scores from other measures that are assumed to represent general cognitive abilities.

Thus, to investigate the validity, the Number Series Study adds a general cognitive ability measure to the PIAAC and PIAAC-L data, namely a measure of (numerical) reasoning. PIAAC-L is a joint research project of three German institutions, GESIS – Leibniz Institute for the Social Sciences, the German Institute for Economic Research (DIW Berlin), and the Leibniz Institute for Educational Trajectories (LIfBi). Funded by the German Federal Ministry of Education and Research, PIAAC-L extends the German PIAAC 2012 assessment by three additional waves, in 2014, 2015, and 2016 (Rammstedt, Martin, Zabal, Carstensen, & Schupp, 2017). One of the overarching goals of PIAAC-L is to bring together different measures of competencies and cognitive skills. The Number Series Study contributes to this goal by providing a measure of reasoning.

The Number Series Study is a research project at the German Institute for International Educational Research (DIPF) within the PIAAC Leibniz Network (PIAAC-LN). PIAAC-LN is directed by GESIS and investigates the acquisition and use of competencies. In addition to the international PIAAC data, PIAAC-LN uses data from the longitudinal PIAAC-L study in Germany. Likewise, the data collection of the Number Series Study is embedded in the PIAAC-L study. This means that the Number Series Study is advised by the PIAAC-L researchers and that the data collection is hosted by the PIAAC-L survey. Selected anchor persons of the PIAAC 2012 sample were assigned again to the PIAAC competence assessments of "Literacy" and "Numeracy" in the second wave of PIAAC-L. A predefined subsample of these respondents participated in the Number Series Study, that is, they were administered a reasoning test in the third wave. Proficiency scores from both assessments, the PIAAC competence assessments and the reasoning test, can now be compared to each other and also in their relation to other variables in the data set, for instance, to relevant background variables. Particularly, the Number Series Study provides data to investigate whether competence measures can be empirically distinguished from measures of reasoning.



This technical report describes the implementation of the Number Series Study for the third wave of PIAAC-L and focuses only on information specific to the Number Series Study. Detailed information on the PIAAC-L study can be found in the PIAAC-L Technical Reports of the three waves (Martin, Zabal, & Rammstedt, 2018; Zabal, Martin, & Rammstedt, 2016, 2017). After giving a brief overview of the goals of the Number Series Study in this chapter, we describe the subsample of PIAAC-L used for the Number Series Study in Chapter 2. Information on the survey instruments can be found in Chapter 3. Chapter 4 provides information about the fieldwork. Information concerning data management and products is given in Chapter 5 and a short conclusion is drawn in Chapter 6.

#### Key Facts: The Number Series Study

- The main objective is to include a reasoning measure into PIAAC-L in order to contribute to one of the objectives of PIAAC-L which is to bring different measures of competency and cognitive skills closer together
- The body responsible is the German Institute for International Educational Research (DIPF)
- Data collection was undertaken as an add-on module in the third PIAAC-L wave in 2016 in cooperation with the PIAAC-L Consortium (GESIS, LfBi, DIW)
- Funded by the Leibniz Association and the Centre for International Student Assessment
- Survey Organization: TNS Infratest (now: Kantar TNS)
- Instruments: Number Series, Confidence Questions, Strategy and Experience Questions
- Interview administration: customized solution by TNS Infratest with interface to CAPI (computer-assisted personal interview), usually administered in the respondent's home and took on average about 10 minutes
- Interview language: German
- Target persons: Anchor persons (participated in PIAAC-L Germany in 2015 and received the computer-based versions of PIAAC competencies)
- Realized sample size: 910
- Data: accessible for scientific purposes as a part of the PIAAC-L scientific use database (ZA5989<sup>1</sup> specifically ZA5989\_NumberSeries\_16) from GESIS Data Archive / Research Data Centre PIAAC (FDZ PIAAC)

---

<sup>1</sup> Version accessible on 14.12.2017: GESIS – Leibniz Institute for the Social Sciences, German Socio-Economic Panel (SOEP) at DIW Berlin & LfBi – Leibniz Institute for Educational Trajectories (2017): PIAAC-Longitudinal (PIAAC-L), Germany. GESIS Data Archive, Cologne. ZA5989 Data file Version 3.0.0, doi: [10.4232/1.12925](https://doi.org/10.4232/1.12925)

---

## 2 The PIAAC-L Sample for the Number Series Study

---

The Number Series Study sample consists of anchor persons from the German PIAAC 2012 sample who were administered computer-based PIAAC instruments in the second wave of PIAAC-L. Anchor persons who did not work on the computer-based versions were not selected for two reasons. For one, the Number Series Study was also administered computer-based, and secondly, the data for the Number Series Study should also include process information. Such process information would be missing for those anchor persons who received the PIAAC assessment as paper versions. The computer-based PIAAC instruments consisted of "Literacy" and "Numeracy". For the Number Series Study which was part of the third wave of PIAAC-L, anchor persons were sampled based on their combination of PIAAC instruments in the second wave (see the design for more details: Zabal et al., 2017). Combinations that were of interest for the Number Series Study were at first those anchor persons who received both instruments, PIAAC Numeracy and Literacy. As these ( $N = 711$ ) did not suffice to realize the target number for this study of 800 cases for data analyses, anchor persons who had received only one of the PIAAC instruments were also selected. These were drawn randomly in equal parts from those who received PIAAC Literacy ( $N = 190$ ) and Numeracy ( $N = 190$ ) first, and then another instrument not relevant to the Number Series Study. Overall, 1,091 anchor persons who participated in 2015 were selected as respondents for the Number Series Study in 2016.

### 3 Instruments

The Number Series Study was implemented as a customized solution by TNS Infratest with an interface to the CAPI (computer-assisted personal interview) software of PIAAC-L. The study consisted of the number series tasks from McArdle and Woodcock (2009; see for further documentation of the number series Fisher, McArdle, McCammon, Sonnega, & Weir, 2013). Questions on confidence and strategy complemented the reasoning tasks; these were especially developed for the purpose of this Number Series Study. For each number series, respondents had to state their confidence concerning the correctness of their solution. After working on all number series and confidence questions, respondents received a few strategy questions.

#### 3.1 Number Series

Fifteen number series in all were part of the Number Series Study. Each number series consisted of at least four and up to seven numbers. One number was missing in each of the first 14 number series, and two numbers were missing in number series 15. All numbers were below 100. Respondents were instructed to respond even in case of uncertainty and to complete as many series as possible. We thus wanted to encourage a response also in case of uncertainty to avoid missing information. However, respondents could state their confidence regarding the correctness of the given response. Respondents started with an example (cf. Figure 1) to familiarize themselves with the computer-based input format and the number series tasks. This exemplary number series is solved correctly if "8" is entered into the response field. Respondents could enter and delete numbers as frequently as they wished. The correct number had to be entered via the number block. Number input fields from zero to nine and a delete-field (cf. Figure 1, "Eingabe löschen") were displayed with each number series. The administered question asks "Which number fits the gap?" (cf. Figure 1, "Welche Zahl gehört in die Leerstelle?").

Welche Zahl gehört in die Leerstelle?

1	2	3
4	5	6
7	8	9
0	Eingabe löschen	

Figure 1: Example of a number series and the input format as it was administered in the field to respondents of the Number Series Study.

The 15 number series were administered in a linear preassigned manner, sorted by difficulty (ascending order). Respondents could self-determine when to navigate to the next number series, but could not navigate backwards. To avoid the respondents' omitting a task unintendedly, they had to confirm their omission first before they received the next number series. On this confirmation page, they still had the choice to navigate back to the task rather than to proceed to the next item. In order to keep to the time planned for the Number Series Study within the context of the PIAAC-L study, a time limit was implemented. The time was counted from the first instruction page until the last number series including all confidence questions. The timeout was set after 16 minutes. If the time expired, the timeout did not interrupt the response process. The system just skipped the remaining number series and led directly to the strategy questions. Respondents were not informed about the time limitation as this was not the intention of the test. Still, they were instructed that there are 15 number series, and a progress bar indicated the task they were currently working on. Respondents did not receive any feedback regarding their task solution.

When respondents navigated to the next item (cf. Figure 1, "Weiter") and had entered a number, or in the case of number series 15 two numbers, they had to state their confidence regarding the correctness of the given response for each number series. If respondents did not enter a number, they did not receive the confidence question, but were asked to confirm their omission. Based on the estimated confidence, we aim at distinguishing serious responses from responses that were guessed using a 4-point Likert Scale (0 = not confident at all (guessing), 1 = not so confident, 2 = confident, 3 = very confident). The confidence question was also implemented in the example task.

### 3.2 Strategy Questions

The strategy questions were administered once to all respondents after they completed the number series questions. These strategy questions were developed for the purpose of this study. They were not part of the time restriction and were also presented if the timeout occurred during the number series. We administered three questions regarding strategic response behavior and one question regarding experience with number series. By administering these strategy questions, we broaden the data and add self-reported behavior during the response process to the behavior for each number series actually recorded by the system. Data from these questions should describe why persons showed a specific response pattern by capturing whether respondents focused rather on speed or on accuracy. The first question focuses on general time usage, the second on double-checking behavior and the third on guessing behavior. The administered questions are presented in Figure 2.

**Beim Bearbeiten der Zahlenaufgaben war es mir wichtig...**

...möglichst zügig voranzukommen.     ...mir ausreichend viel Zeit zu nehmen.

---

...sofort weiter zu klicken sobald ich die Lösung eingetragen habe.     ...zur Sicherheit noch einmal über die Lösung nachzudenken, sobald ich die Lösung eingetragen habe.

---

...zu schätzen, wenn ich die Lösung nicht gleich gefunden habe, um nicht zu viel Zeit zu verlieren.     ...die Aufgabe exakt zu lösen auch wenn es so länger gedauert hat.

**Haben Sie solche Zahlenaufgaben schon einmal bearbeitet?**

Nein, noch nie       Ja, einmal       Ja, mehr als einmal

Figure 2: Strategy and experience questions as they were administered in the field to the respondents of the Number Series Study.

As shown in Figure 2, we did not label each response category but provided only statements for the two extreme categories. The codebook also provides labels for the categories in between. See Table 1 for the back translation of the German original text into English.

Table 1: Back translation of the German original text from Figure 2 into English.

Variable names*	Translations into English	
	<b>While working on the number series, it was important to me to...</b>	
q_strategy1_16	...proceed as quickly as possible.	...take as much time as I need.
q_strategy2_16	...proceed immediately after having entered the solution.	...rethink the solution, after having entered the number, to be on the safe side.
q_strategy3_16	...make a rough guess, if I could not solve it straightaway, in order not to lose too much time.	...solve the task exactly, even if it took more time this way.
q_experience_16	<b>Did you ever do number series before?</b>	
	No, never	Yes, once      Yes, several times

Notes. \*variable names in the codebook and the SUF.

## 4 Fieldwork

---

The Number Series Study is embedded in the third wave of the PIAAC-L study (see Martin et al., 2018). Before they started with the number series, the respondents were informed that DIPF was responsible for this additional test. For the number series, the laptop was modified. As the laptops were foldable, the screen was turned down and placed in a horizontal position on the table. The respondents used a pen and the numerical entry panel on the screen, as shown in Figure 1, to enter the numbers. This entry format was not only considered to be more convenient, but had the advantage that the keyboard of the laptop was covered by this turned down position. As a consequence, all respondents entered the numbers via the numerical entry panel making response times more comparable.

Respondents completed this part of the interview autonomously and the interviewer was not supposed to provide any guidance. Interviewers received an additional fee for administering this part of the interview.

## 5 Data Management, Data Products, and Documentation

Data cleaning occurred in three different steps with three different purposes. Technical issues were traced and documented in a first step by the survey organization, TNS Infratest. As this study is part of the PIAAC-L study, issues with identifiers were checked and resolved by GESIS. Final plausibility checks with respect to the Number Series data were conducted by the DIPF team.

The final sample of the Number Series Study is composed as follows: out of the 1,091 respondents (anchor persons) drawn from the second wave of PIAAC-L for the Number Series Study, 976 participated in the PIAAC-L study in the third wave. Out of those 976 respondents,  $n = 27$  (2.8 %) respondents refused to participate in the Number Series Study after they completed the PIAAC-L instruments. Technical issues led to missing, thus not recorded, administrations for  $n = 15$  (1.5 %) for the Number Series Study. A few,  $n = 13$  (1.3 %), respondents had implausible data and  $n = 9$  (0.9 %) respondents worked on the number series possibly twice due to technical issues. Because this called comparability into question, these cases were flagged by the survey organization based on data and interviewer comments and removed at DIPF from the final data set. Two further respondents (0.2 %) were finally removed at DIPF as data for the number series was missing for unknown reasons. The final SUF contains the remaining 910 cases. Respondents of the Number Series Study were on average  $M = 43.47$  years old ( $SD = 13.62$ ,  $Min = 20$ ,  $Max = 69$ ) and about half of them ( $n = 452$ ) were male (female:  $n = 458$ ).

These 910 respondents worked on the Number Series, including the instruction and the confidence questions, on average  $M = 9.49$  minutes ( $SD = 5.3$ ,  $Min = 3.9$ ,  $Max = 19.6$ ). The strategy questions were not part of this time measurement and took on average  $M = 41.35$  seconds ( $SD = 22.38$ ). Four respondents did not receive the strategy questions due to technical irregularities. Table 2 contains some further information about the response behavior of the respondents across all number series. Only seven respondents did not reach all number series due to timeout. Omission attempts were always higher than the actual omissions, supporting the conjecture that at least some respondents returned to the number series and gave a response. Only a few respondents actually omitted number series. The number of omissions increased with item position and increasing difficulty. Deletions of once given responses were seldom and occurred especially on the last number series. Respondents who omitted or deleted a given answer did not receive the confidence questions. This also applied to those who left number series 15 without providing a complete answer, either because they filled in only one gap or they filled in both gaps but deleted one number again. These incomplete responses were treated as incorrect.

Table 2: Missing responses across all number series

Number Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
not reached (-6)	0	0	0	0	0	0	0	0	0	0	1	2	2	2	7
omission attempt	17	6	1	5	6	29	2	6	12	33	21	17	28	31	139
omitted (8)	4	1	1	1	3	15	2	3	8	23	18	10	23	26	59
deleted response (7)	0	0	0	0	0	1	0	0	0	1	1	1	0	1	33

Notes. The codes in brackets correspond to the values of the score variable within the data set. Omission attempts were not separately coded. See the codebook for documentation.



In order to link the Number Series Study to other PIAAC-L data, three person identifiers are available in the data set of the Number Series Study: permanent person ID (pnrfestid), household ID (hnrld), and sequential ID (seqid). With the sequential ID, which refers to anchor persons from PIAAC 2012, data from respondents of the Number Series Study can be linked to the PIAAC 2012 data set and also to data from the first, second, and third wave of PIAAC-L.

The Number Series data set contains six variables for number series 1-14 and nine variables for number series 15, because two gaps had to be filled in the last number series. The following variables are available in the SUF:

- One score variable per number series indicates whether the number series was solved correctly or incorrectly, with two additional codes if no response was entered at all (omitted) or the respondent deleted a given response. This way, persons who solved the item incorrectly by entering an incorrect response can be distinguished from persons who did not even guess. It is up to the data user to treat deleted responses and omissions as incorrect.
- One variable per number series describes the respondents' confidence of correctness of a given response.
- Four variables per number series describe the response process. They include the time spent on the task, after how many seconds the first and last attempt was made after the number series was displayed, and a count variable for the number of attempts. The last three variables were recorded for both gaps of the number series 15. These time variables were selected as they describe significant moments during the response process. Compared to the overall time spent on the task, the timing of the attempts describes in more detail when the response was actually given. The number of attempts indicates how intensively the respondent interacted with the task and might relate also to the time spent on the task.
- For the strategy and experience questions, one variable per question is available as well as one variable indicating the time spent on these questions.

Data from the Number Series Study is part of the PIAAC-L database, which is available for scientific use via the Research Data Centre PIAAC (RDC/PIAAC FDZ/PIAAC). In order to obtain access to the data, it is necessary to register and sign the specific PIAAC-L Data Use Agreement. This agreement gives access to data from all three waves of PIAAC-L. Data from the Number Series Study can be linked via the identifiers to the PIAAC-L data.

Data from the Number Series Study can be found in a separate data set (ZA5989\_NumberSeries\_16). It contains all valid interviews from the Number Series Study (Units: 910 respondents; Content: 104 variables).

An extensive codebook (in English) is available as well as the survey organization's fieldwork report (in German only; Steinacker, Wolfert, & Thümmel, 2017).

---

## 6 Conclusion

---

Most of the respondents (anchor persons) from the second wave of PIAAC-L that were selected for participation in the Number Series Study participated in the third wave in the Number Series Study. Data from 83.4 % of the initially drawn respondents can be used for analyses. Most of the respondents completed the Number Series within the suggested time frame and rates of missing data are rather low. The Number Series Study, as part of the PIAAC-L 2016 data collection, extends and enriches the PIAAC-L data and is accessible to the scientific community.

## References

---

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education [AERA, APA, & NCME] (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Fisher, G. G., McArdle, J. J., McCammon, R. J., Sonnega, A., & Weir, D. R. (2013). New Measures of Fluid Intelligence in the HRS, HRS Documentation Report DR-027. Survey Research Center, University of Michigan.
- Martin, S., Zabal, A., & Rammstedt, B. (2018). *PIAAC-L data collection 2016: Technical report*. Manuscript in preparation.
- McArdle, J. J., & Woodcock, R. W. (2009). *Adaptive testing with selected Woodcock cognitive tests* [Powerpoint slides]. Charlottesville, VA: Longitudinal Research Institute.
- OECD (2016). *The Survey of Adult Skills: Reader's Companion, Second Edition*. OECD Skills Studies, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264258075-en>
- Rammstedt, B., Martin, S., Zabal, A., Carstensen, C., & Schupp, J. (2017). The PIAAC longitudinal study in Germany: rationale and design. *Large-scale Assessments in Education*, 5, 1-11. <https://doi.org/10.1186/s40536-017-0040-z>
- Steinacker, G., Wolfert, S., & Thümmel, K. (2017). *Durchführung der 3. Erhebungswelle von PIAAC-L (Kooperative längsschnittliche Weiterverfolgung der PIAAC-Studie in Deutschland). Feldbericht zur Erhebung 2016*. GESIS Papers 2017|27. Cologne: GESIS. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-54976-6>
- Zabal, A., Martin, S., & Rammstedt, B. (2016). *PIAAC-L data collection 2014: Technical report*. GESIS Papers 2016|17. GESIS: Cologne. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-49665-1>
- Zabal, A., Martin, S., & Rammstedt, B. (2017). *PIAAC-L Data Collection 2015: Technical report*. GESIS Papers 2017|29. Cologne: GESIS. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-55155-8>