The PIAAC Variable Finder: User's Guide

Nate Breznau, breznau.nate@gmail.com,

German Institute for Adult Education – Leibniz Centre for Lifelong Learning

Version 0.2 24.07.2025

Guide for users of the PIAAC Variable Finder

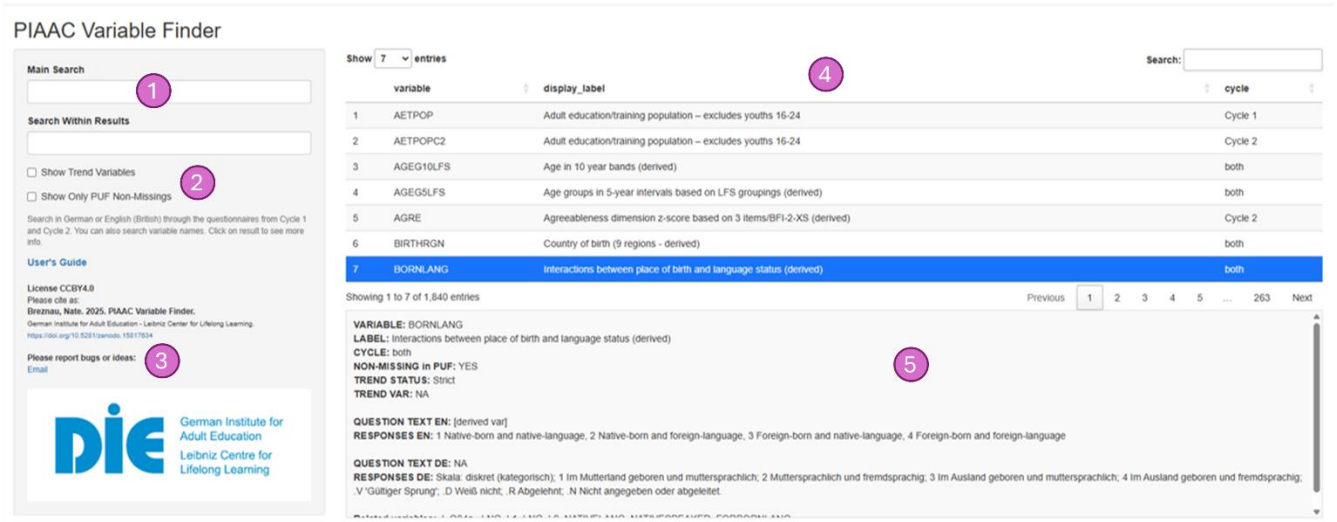
Working paper: Breznau, Nate. 2025. "The PIAAC Variable Finder: An Interactive Shiny App for Cleaning, Interpreting and Analyzing Programme for the International Assessment of Adult Competencies Data." SocArXiv. doi:10.31235/osf.io/xp4ja_v1.

Web-based version: https://nate-breznau.shinyapps.io/PIAAC_Variable_Finder/

Locally deployable folder (as zip file): https://github.com/nbreznau/PIAAC_Variable_Finder/blob/main/Shiny.zip

Development workflow on Github: https://github.com/nbreznau/PIAAC_VET_basics

User Interface – Quick Start



1. **Search**
    a. "Main Search" will filter on exact text strings. It is not a Boolean search.
    b. "Search Within Results". To get sub filtering (like a Boolean "AND") use this bar.
2. **Subsetting Options**
    a. "Show Trend Variables". When ticked, only variables that are a 'strict' or 'soft' trend, or derived and repeated across both Cycles will be shown.
    b. "Show Only PUF Non-Missings". When ticked, only variables that have at least one country-cycle set of non-missing values samples will be shown (for Germany and the United Kingdom).
3. **Links**
    a. User's Guide links to the most current version of this document
    b. License CCBY4.0 means that all materials are freely usable for non-commercial purposes with proper citation
    c. Suggested citation is: Breznau, Nate. 2025. PIAAC Variable Finder. German Institute for Adult Education - Leibniz Center for Lifelong Learning. https://doi.org/10.5281/zenodo.15817634
    d. A link to email Nate Breznau with bug reporting, suggestions or comments
4. **Search Results**
    a. Interactive. Click (or activate with touch screen) on a given result to highlight it and then see the appropriate metadata below.
    b. Seven per page. Click on the numbers below this window to advance through the pages of results. All results are shown at first, but all searching and subsetting is dynamic, so it updates as you type or tick.
    c. Includes the main information: variable name as it is in the PIAAC data, label taken from PIAAC and which cycle it is in
5. **Metadata**
    a. Trend. Includes the "STATUS" if it is 'Strict' (identical between cycles), 'Soft' (comparable, but different wording and/or answer categories), 'Derived' (a constructed variable exists in both Cycles, theoretically a trend, but measurement can differ), 'NA' not a trend, or 'Cy 1' or 'Cy 2' only.
    b. Verbatim question text in English and German (except for derived variables)
    c. Related variables. Variables that ask related questions, are used to construct the variable, or link trends

Detailed Description

The user interface (UI) contains three primary regions. The left-hand sidebar has two search fields (Figure 1.1). The Main Search filters the entire database directly. It is not Boolean, so words and phrases must be exact. As of version 1.0, most fields have both English and German meta data. The app does not engage in any translation however, so in some cases the ostensibly same word in both languages could yield different results. The Search Within Results field searches within those results already showing.

Next in the sidebar are two check boxes that default as unticked (Figure 1.2). If the user ticks the Show Trend Variables box, it will filter the results to include only strict and soft trend variables. Corresponding to this is the "TREND STATUS" variable in Figure 1.5 which can have "NA" which means 'not applicable' and indicates no trend, "Derived" which indicates it is most likely a trend variable but the variable from the other Cycle has a different name (usually with "C2" at the end, "Cy 1 only" or "Cy 2 only" which is self-evidence, and "soft" or "strict" indicating face (roughly speaking) or both face and content validity (strictly speaking) in the measure across the two Cycles. If the user ticks the Show Only PUF Non-Missings box, it will only show variables that have non-missing data in at least one cycle and at least one country (Germany or the United Kingdom). In other words, when this box is not ticked it will show many variables that exist in PIAAC but are all missing values in the PUFs. Corresponding to this variable in Figure 1.5 is the field for "NON-MISSING in PUF:", which will populate with the value "No" for every variable when the box is ticked.

Figure 1.3 is a link to the User's Guide which contains the same basic information relayed here, but with a bullet point structure and more details as part of the app's technical documentation. Directly below this is the preferred citation. The DOI for the app works through Zenodo. This is a free to use service that has a Github plugin. It automatically generates a DOI and keeps up with newer versions of the repository supporting the app. Zenodo works stand alone and with many other workflow and repository services. Thus, Zenodo is an ideal open science tool.

Figure 1.4 is the main results panel in the center-right region of the app. It displays the first 7 variable results in alphabetical order (the order of the dataframe). I elected to display 7 per page, because on most devices this seemed to maximize space allowing the other two panels to be viewable on most single monitor screens. These features are all easily customizable. Once a user is familiar with Shiny, it is incredibly easy to use, especially by asking Gen AI. My point is that user's are encouraged to customize their experience. In addition, feedback is welcome, therefore the final line at the bottom of Figure 1.3 (the side panel) is a link to my email to suggest modifications and report any bugs.

In the list for Figure 1.4 the user sees the variable name verbatim as it is in the PIAAC data in the column "variable". Then under "display_label" is the value for variable_label_en which is the metadata variable for the English label of the variable taken mostly from the PIAAC's own labelling. There are exceptions because some countries have their own questions, especially when it comes to the education system. For example, the UK as a variable B_Q01a3UK from the question, "Can you indicate which level in our national education system corresponds most closely with the level of this qualification?". In Germany a similarly unique variable B_Q01aDE1 is measured with the question, "Welchen höchsten allgemeinbildenden Schulabschluss haben Sie? Bitte sagen Sie es mir anhand dieser Liste." These completely country-unique variables as a rule have the country's two-digit alpha International Standards Organization (iso2c) classification in the variable name. Because these variables do not have labels, I use the question wording from the questionnaires as replacement – I programed it this way, so that if there is no label, then the question wording appears. The last column "cycle" lists from which Cycle the variable derives, or if it is in "both".

Figure 1.5 is a more detailed breakdown of the highlighted variable. This appears blank when the app is launched, and is populated with values only after the user selects (clicks or taps) one of the seven rows in the list so that it is highlighted, i.e., active. Then nearly all available metadata for that variable appears below. Table 1 lists all variables in the metadata that I extracted from the PIAAC documentation and the PIAAC data itself. This provides an explanation for the metadata results in this section.

Table 1. Metadata Variables in the Dataframe Behind the PIAAC Variable Finder

meta variable	purpose	values	displayed?
variable	variable name in PIAAC data	verbatim from PUF files	yes
question_text_de	Question wording German	Verbatim from questionnaire[a]	yes
question_text_en	Question wording English	Verbatim from questionnaire[a]	yes
responses_de	Response wording German	Verbatim from questionnaire[a]	yes
responses_en	Response wording German	Verbatim from questionnaire[a]	yes
generic_label_de	Variable label German	From PIAAC documentation	no
generic_label_en	Variable label English	From PIAAC documentation	yes
constructed_vars	Rules for variable construction	From PIAAC documentation	no
ref_variables	A list of related variables, including those used to construct this variable[b]	If/then rules	no
cycle	Cycle identifier	• Cy 1 • Cy 2 • both	yes
trend	Identifies a trend variable	• Strict • Soft • Derived [c] • Cy1 only • Cy2 only • NA (not a trend)	yes
trend_var	Identifies trend pair	If different between Cycles, the name of the corresponding trend variable	yes
soft_trend_explanation	Differences between Cycles	PIAAC's documented explanation	yes, if trend = "Soft"
c_vars	Related variables	Compiles base variables for derived measures, and similar variables	yes
notin	If all missing in PUF for DE & UK Cycle 1	1 = Yes	no
notin2	If all missing in PUR for DE & UK Cycle 2	1 = Yes	no
none	If missing for both Cycles and countries	• Yes • No	yes

[a]Except for derived variables which use the PIAAC documentation descriptive wording

[b]There uncertainty here (room for development) because of the diverse sources for this information

[c]"Derived" indicates a constructed variable that exists in both Cycles, thus theoretically a trend, but measurement can differ. Should be investigated on a case-by-case basis for comparison