

# Package ‘educabR’

April 3, 2026

**Title** Download and Process Brazilian Education Data from INEP

**Version** 0.9.0

**Description** Download and process public education data from INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). Provides functions to access microdata from the School Census (Censo Escolar), ENEM (Exame Nacional do Ensino Médio), SAEB (Sistema de Avaliação da Educação Básica), Higher Education Census (Censo da Educação Superior), ENADE (Exame Nacional de Desempenho dos Estudantes), ENCCEJA (Exame Nacional para Certificação de Competências de Jovens e Adultos), IDD (Indicador de Diferença entre os Desempenhos Observado e Esperado), CPC (Conceito Preliminar de Curso), IGC (Índice Geral de Cursos), CAPES graduate education data, FUNDEB (Fundo de Manutencao e Desenvolvimento da Educacao Basica), IDEB (Índice de Desenvolvimento da Educação Básica), and other educational datasets. Returns data in tidy format ready for analysis. Data source: INEP Open Data Portal <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos>>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**RoxygenNote** 7.3.3

**URL** <https://github.com/SidneyBissoli/educabR>,  
<https://sidneybissoli.github.io/educabR/>

**BugReports** <https://github.com/SidneyBissoli/educabR/issues>

**Depends** R (>= 4.1.0)

**Imports** cli, dplyr, httr2, purrr, readr, rlang, stringr, tidyr, tools

**Suggests** ggplot2, jsonlite, knitr, readxl, rmarkdown, testthat (>= 3.0.0), tibble, withr

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Sidney da Silva Pereira Bissoli [aut, cre] (ORCID:  
<https://orcid.org/0009-0001-0442-3700>)

**Maintainer** Sidney da Silva Pereira Bissoli <sbissoli76@gmail.com>

**Repository** CRAN

**Date/Publication** 2026-04-03 07:00:08 UTC

## Contents

available_years . . . . .	2
clear_cache . . . . .	3
enem_summary . . . . .	4
get_cache_dir . . . . .	5
get_capes . . . . .	5
get_censo_escolar . . . . .	7
get_censo_superior . . . . .	8
get_cpc . . . . .	10
get_enade . . . . .	11
get_encceja . . . . .	13
get_enem . . . . .	14
get_enem_escola . . . . .	16
get_enem_itens . . . . .	17
get_fundeb_distribution . . . . .	18
get_fundeb_enrollment . . . . .	20
get_idd . . . . .	21
get_ideb . . . . .	23
get_ideb_series . . . . .	24
get_igc . . . . .	25
get_saeb . . . . .	26
list_cache . . . . .	28
list_censo_files . . . . .	29
list_censo_superior_files . . . . .	29
list_ideb_available . . . . .	30
set_cache_dir . . . . .	31
<b>Index</b>	<b>32</b>

---

available_years	<i>Check available years for a dataset</i>
-----------------	--

---

## Description

Returns the years available for a given dataset. On the first call in a session, queries the data source to discover which years are actually available (requires internet). Results are cached for the session. Falls back to a known list if discovery fails.

**Usage**

```
available_years(dataset)
```

**Arguments**

dataset            The dataset name.

**Value**

An integer vector of available years.

**Examples**

```
## Not run:  
available_years("enem")  
available_years("enade")  
available_years("fundeb_enrollment")  
  
## End(Not run)
```

---

clear_cache	<i>Clear the educabR cache</i>
-------------	--------------------------------

---

**Description**

Removes all cached files from the educabR cache directory.

**Usage**

```
clear_cache(dataset = NULL)
```

**Arguments**

dataset            Optional. A character string specifying which dataset cache to clear. If NULL, clears all caches.

**Value**

Invisibly returns TRUE if successful.

**See Also**

Other cache functions: [get\\_cache\\_dir\(\)](#), [list\\_cache\(\)](#), [set\\_cache\\_dir\(\)](#)

## Examples

```
## Not run:  
# clear all cached data  
clear_cache()  
  
# clear only ENEM cache  
clear_cache("enem")  
  
## End(Not run)
```

---

enem\_summary

*Summary statistics for ENEM scores*

---

## Description

Calculates summary statistics for ENEM scores, optionally grouped by demographic variables.

## Usage

```
enem_summary(data, by = NULL)
```

## Arguments

**data** A tibble with ENEM data (from [get\\_enem\(\)](#)).  
**by** Optional grouping variable(s) as character vector.

## Value

A tibble with summary statistics for each score area.

## See Also

Other ENEM functions: [get\\_enem\(\)](#), [get\\_enem\\_escola\(\)](#), [get\\_enem\\_itens\(\)](#)

## Examples

```
## Not run:  
enem <- get_enem(2023, n_max = 10000)  
  
# overall summary  
enem_summary(enem)  
  
# summary by sex  
enem_summary(enem, by = "tp_sexo")  
  
## End(Not run)
```

---

get_cache_dir	<i>Get the current cache directory</i>
---------------	--

---

**Description**

Returns the current cache directory used by educabR.

**Usage**

```
get_cache_dir()
```

**Value**

A character string with the path to the cache directory.

**See Also**

Other cache functions: [clear\\_cache\(\)](#), [list\\_cache\(\)](#), [set\\_cache\\_dir\(\)](#)

**Examples**

```
get_cache_dir()
```

---

get_capes	<i>Get CAPES graduate education data</i>
-----------	--

---

**Description**

Downloads and processes data from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) on Brazilian graduate programs (stricto sensu). Data is retrieved from the CAPES Open Data Portal via the CKAN API.

**Usage**

```
get_capes(  
  year,  
  type = c("programas", "discentes", "docentes", "cursos", "catalogo"),  
  n_max = Inf,  
  keep_file = TRUE,  
  quiet = FALSE  
)
```

### Arguments

year	The year of the data (2013-2024).
type	The type of data to download. One of: <ul style="list-style-type: none"><li>• "programas": Graduate programs</li><li>• "discentes": Students</li><li>• "docentes": Faculty</li><li>• "cursos": Courses</li><li>• "catalogo": Theses and dissertations catalog</li></ul>
n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_file	Logical. If TRUE, keeps the downloaded file in cache. Default is TRUE.
quiet	Logical. If TRUE, suppresses progress messages.

### Details

CAPES is the federal agency responsible for evaluating and regulating graduate programs in Brazil. The data covers stricto sensu programs (master's and doctoral).

The data types include:

- **programas**: Program identifiers, area, evaluation scores
- **discentes**: Student enrollment, demographics, scholarship status
- **docentes**: Faculty information, qualifications, employment
- **cursos**: Course details, modality, start dates
- **catalogo**: Catalog of theses and dissertations

### Important notes:

- Data is sourced from the CAPES Open Data Portal (CKAN), not INEP.
- Files are large CSV files. Downloading may take several minutes.
- Column names are standardized to lowercase with underscores.
- Internet connection is required to discover download URLs via the CKAN API before downloading.

### Value

A tibble with CAPES data in tidy format.

### Data source

<https://dadosabertos.capes.gov.br>

## Examples

```
## Not run:  
# get graduate programs for 2023  
programas <- get_capes(2023, type = "programas")  
  
# get student data for 2022 with limited rows  
discentes <- get_capes(2022, type = "discentes", n_max = 1000)  
  
## End(Not run)
```

---

get\_censo\_escolar      *Get School Census (Censo Escolar) data*

---

## Description

Downloads and processes microdata from the Brazilian School Census (Censo Escolar), conducted annually by INEP. Returns school-level data with information about infrastructure, location, and administrative details.

## Usage

```
get_censo_escolar(year, uf = NULL, n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

## Arguments

year	The year of the census (1995-2024).
uf	Optional. Filter by state (UF code or abbreviation).
n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

## Details

The School Census is the main statistical survey on basic education in Brazil. It collects data from all public and private schools offering basic education (early childhood, elementary, and high school).

### Important notes:

- The microdata contains one row per school (~217,000 schools in 2023).
- Column names are standardized to lowercase with underscores.
- Use the `uf` parameter to filter by state for faster processing.

## Value

A tibble with school data in tidy format.

## Data dictionary

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-escolar>

## See Also

Other School Census functions: `list_censo_files()`

## Examples

```
## Not run:
# get schools data for 2023
escolas <- get_censo_escolar(2023)

# get schools from Sao Paulo state only
escolas_sp <- get_censo_escolar(2023, uf = "SP")

# read only first 1000 rows for exploration
escolas_sample <- get_censo_escolar(2023, n_max = 1000)

## End(Not run)
```

---

get\_censo\_superior      *Get Higher Education Census (Censo da Educação Superior) data*

---

## Description

Downloads and processes microdata from the Brazilian Higher Education Census (Censo da Educação Superior), conducted annually by INEP. Returns data on institutions, courses, students, or faculty.

## Usage

```
get_censo_superior(
  year,
  type = c("ies", "cursos", "alunos", "docentes"),
  uf = NULL,
  n_max = Inf,
  keep_zip = TRUE,
  quiet = FALSE
)
```

## Arguments

`year`                    The year of the census (2009-2024).

`type`                    Type of data to load. Options:

- "ies": Higher education institutions (default)

	<ul style="list-style-type: none"> <li>• "cursos": Undergraduate courses</li> <li>• "alunos": Student enrollment</li> <li>• "docentes": Faculty/professors</li> </ul>
uf	Optional. Filter by state (UF code or abbreviation).
n_max	Maximum number of rows to read. Default is Inf (all rows). Consider using a smaller value for exploration.
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

### Details

The Higher Education Census is the most comprehensive statistical survey on higher education institutions (HEIs) in Brazil. It collects data from all HEIs offering undergraduate and graduate programs.

#### Data types:

- "ies": One row per institution — administrative data, location, academic organization, funding type.
- "cursos": One row per undergraduate course — area of study, modality (in-person/distance), enrollment counts.
- "alunos": One row per student enrollment — demographics, program, admission type, enrollment status.
- "docentes": One row per faculty member — education level, employment type, teaching regime.

#### Important notes:

- Student files ("alunos") can be very large (several GB). Use n\_max to read a sample first.
- Column names are standardized to lowercase with underscores.
- Use the uf parameter to filter by state for faster processing.

### Value

A tibble with Higher Education Census microdata in tidy format.

### Data dictionary

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/acao-a-informacao/dados-abertos/microdados/cento-da-educacao-superior>

### See Also

Other Higher Education Census functions: [list\\_censo\\_superior\\_files\(\)](#)

**Examples**

```
## Not run:
# get institution data for 2023
ies <- get_censo_superior(2023)

# get course data for Sao Paulo
cursos_sp <- get_censo_superior(2023, type = "cursos", uf = "SP")

# get a sample of student data
alunos <- get_censo_superior(2023, type = "alunos", n_max = 10000)

# get faculty data
docentes <- get_censo_superior(2023, type = "docentes")

## End(Not run)
```

---

get\_cpc

*Get CPC (Conceito Preliminar de Curso) data*


---

**Description**

Downloads and processes CPC data from INEP. The CPC is a quality indicator for undergraduate courses in Brazil, composed of ENADE scores, IDD, faculty qualifications, pedagogical resources, and other institutional factors.

**Usage**

```
get_cpc(year, n_max = Inf, keep_file = TRUE, quiet = FALSE)
```

**Arguments**

year	The year of the indicator (2007-2019, 2021-2023). Note: there is no 2020 edition. Years 2004-2006 used a different indicator ("Conceito Enade").
n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_file	Logical. If TRUE, keeps the downloaded file in cache. Default is TRUE.
quiet	Logical. If TRUE, suppresses progress messages.

**Details**

CPC is calculated by INEP as part of the higher education quality assessment system (SINAES). It serves as a preliminary indicator used to determine which courses require on-site evaluation.

The data includes:

- Course and institution identifiers
- CPC scores (continuous and categorical/faixa)
- Component scores (ENADE, IDD, faculty, infrastructure, etc.)

- Number of students evaluated

**Important notes:**

- CPC follows ENADE's rotating cycle of course areas, so each year covers a specific set of fields.
- There is no 2020 edition (COVID-19 suspension).
- Column names are standardized to lowercase with underscores.
- Files are in Excel format (xls/xlsx), not CSV.

**Value**

A tibble with CPC data in tidy format.

**Data dictionary**

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/indicadores-de-qualidade-da-educacao-s>

**See Also**

Other CPC/IGC functions: [get\\_igc\(\)](#)

**Examples**

```
## Not run:  
# get CPC data for 2023  
cpc <- get_cpc(2023)  
  
# get CPC data for 2021 with limited rows  
cpc_2021 <- get_cpc(2021, n_max = 1000)  
  
## End(Not run)
```

---

get\_enade

*Get ENADE (Exame Nacional de Desempenho dos Estudantes) data*

---

**Description**

Downloads and processes microdata from ENADE, the Brazilian National Student Performance Exam. ENADE evaluates the performance of undergraduate students in higher education.

**Usage**

```
get_enade(year, n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

### Arguments

year	The year of the exam (2004-2024).
n_max	Maximum number of rows to read. Default is Inf (all rows). Consider using a smaller value for exploration.
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

### Details

ENADE is conducted annually by INEP and evaluates undergraduate students nearing the end of their programs. Each year, a different set of course areas is assessed on a rotating cycle (typically every 3 years per area).

The microdata includes:

- Student performance scores (general and specific knowledge)
- Socioeconomic questionnaire responses
- Course and institution identifiers

### Important notes:

- ENADE files can be large (several hundred MB for recent years).
- Use n\_max to read a sample first for exploration.
- Column names are standardized to lowercase with underscores.
- Not all course areas are assessed every year due to the rotating cycle.

### Value

A tibble with ENADE microdata in tidy format.

### Data dictionary

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enade>

### Examples

```
## Not run:  
# get ENADE data for 2023  
enade <- get_enade(2023, n_max = 10000)  
  
# get full dataset for 2021  
enade_2021 <- get_enade(2021)  
  
## End(Not run)
```

---

get_encceja	<i>Get ENCCEJA (Exame Nacional para Certificação de Competências de Jovens e Adultos) data</i>
-------------	--

---

### Description

Downloads and processes microdata from ENCCEJA, the Brazilian National Exam for Youth and Adult Education Certification. ENCCEJA assesses competencies of young people and adults who did not complete basic education at the regular age.

### Usage

```
get_encceja(year, n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

### Arguments

year	The year of the exam (2014-2024).
n_max	Maximum number of rows to read. Default is Inf (all rows). Consider using a smaller value for exploration.
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

### Details

ENCCEJA is conducted by INEP and provides certification for elementary and high school equivalency for youth and adults (EJA). The exam covers four knowledge areas:

- Natural Sciences (Ciências Naturais)
- Mathematics (Matemática)
- Portuguese Language (Língua Portuguesa)
- Social Sciences (Ciências Humanas)

### Important notes:

- ENCCEJA files can be large (several hundred MB).
- Use n\_max to read a sample first for exploration.
- Column names are standardized to lowercase with underscores.

### Value

A tibble with ENCCEJA microdata in tidy format.

### Data dictionary

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/encceja>

**Examples**

```
## Not run:
# get ENCEJA data for 2023
encceja <- get_encceja(2023, n_max = 10000)

# get full dataset for 2022
encceja_2022 <- get_encceja(2022)

## End(Not run)
```

---

get\_enem

*Get ENEM (Exame Nacional do Ensino Médio) data*


---

**Description**

Downloads and processes microdata from ENEM, the Brazilian National High School Exam. ENEM is used for university admissions and as a high school equivalency exam.

**Usage**

```
get_enem(
  year,
  type = "participantes",
  n_max = Inf,
  keep_zip = TRUE,
  quiet = FALSE
)
```

**Arguments**

year	The year of the exam (1998-2024).
type	Type of data to load. Only used for ENEM 2024+, where microdata is split into separate files. Options: "participantes" (demographics and socioeconomic data, default), "resultados" (scores). Ignored for years before 2024 (single file with all data).
n_max	Maximum number of rows to read. Default is Inf (all rows). Consider using a smaller value for exploration, as ENEM files contain millions of rows.
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

**Details**

ENEM is conducted annually by INEP and is the largest exam in Brazil, with millions of participants. The microdata includes:

- Participant demographics (age, sex, race, etc.)

- Socioeconomic questionnaire responses
- Scores for each test area
- Essay scores
- School information (when applicable)

**Important notes:**

- ENEM files are very large (several GB when extracted).
- Use `n_max` to read a sample first for exploration.
- Column names are standardized to lowercase with underscores.
- Score variables start with `nu_not_a_` prefix.
- From 2024 onwards, INEP split the microdata into separate files. Use the `type` parameter to choose which file to load.

**Value**

A tibble with the ENEM microdata in tidy format.

**Data dictionary**

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>

**See Also**

Other ENEM functions: [enem\\_summary\(\)](#), [get\\_enem\\_escola\(\)](#), [get\\_enem\\_itens\(\)](#)

**Examples**

```
## Not run:
# get a sample of 10000 rows for exploration
enem_sample <- get_enem(2023, n_max = 10000)

# get full data (warning: large file)
enem_2023 <- get_enem(2023)

# ENEM 2024+: choose data type
participantes <- get_enem(2024, type = "participantes", n_max = 1000)
resultados <- get_enem(2024, type = "resultados", n_max = 1000)

## End(Not run)
```

---

get_enem_escola	<i>Get ENEM por Escola (ENEM by School) data</i>
-----------------	--

---

### Description

Downloads and processes ENEM results aggregated by school. This dataset contains average ENEM scores, participation rates, and other indicators for each school in Brazil.

### Usage

```
get_enem_escola(n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

### Arguments

n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

### Details

ENEM por Escola is a single bundled dataset covering years 2005 to 2015. It was discontinued by INEP after 2015 and no per-year files exist.

The data includes:

- School identification (code, name, municipality, state)
- Average ENEM scores by knowledge area
- Number of participants and participation rates
- School-level indicators

### Important notes:

- This is a single file covering all years (2005-2015), not per-year.
- Column names are standardized to lowercase with underscores.
- Data was discontinued after 2015.

### Value

A tibble with ENEM by School data in tidy format.

### Data dictionary

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem-por-escola>

### See Also

Other ENEM functions: [enem\\_summary\(\)](#), [get\\_enem\(\)](#), [get\\_enem\\_itens\(\)](#)

## Examples

```
## Not run:  
# get all ENEM by School data (2005-2015)  
enem_escola <- get_enem_escola()  
  
# read only first 1000 rows for exploration  
enem_escola_sample <- get_enem_escola(n_max = 1000)  
  
## End(Not run)
```

---

get_enem_itens	<i>Get ENEM item response data</i>
----------------	------------------------------------

---

## Description

Downloads and processes ENEM item response (gabarito) data, which contains detailed information about each question.

## Usage

```
get_enem_itens(year, n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

## Arguments

year	The year of the exam (1998-2024).
n_max	Maximum number of rows to read.
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

## Value

A tibble with item response data.

## See Also

Other ENEM functions: [enem\\_summary\(\)](#), [get\\_enem\(\)](#), [get\\_enem\\_escola\(\)](#)

## Examples

```
## Not run:  
# get item data for 2023  
itens <- get_enem_itens(2023)  
  
## End(Not run)
```

---

```
get_fundeb_distribution
```

*Get FUNDEB distribution data*

---

### Description

Downloads and processes FUNDEB resource distribution data from STN (Secretaria do Tesouro Nacional). Each year's Excel file contains multiple sheets with monthly transfer data by state/municipality, broken down by funding source.

### Usage

```
get_fundeb_distribution(
    year,
    uf = NULL,
    source = NULL,
    destination = NULL,
    n_max = Inf,
    keep_file = TRUE,
    quiet = FALSE
)
```

### Arguments

year	The year of the data (2007-2026).
uf	Optional. A UF code (e.g., "SP", "RJ") to filter by state. Default is NULL (all states).
source	Optional. The funding source to filter by. One of: "FPE", "FPM", "IPI", "ITR", "VAAF", "VAAT", "VAAR", "ICMS", "IPVA", "ITCMD". Default is NULL (all sources).
destination	Optional. The transfer destination. One of: <ul style="list-style-type: none"> <li>"uf": Transfers to states and the Federal District</li> <li>"municipio": Transfers to municipalities</li> </ul> Default is NULL (both).
n_max	Maximum number of rows to return. Default is Inf (all rows).
keep_file	Logical. If TRUE, keeps the downloaded file in cache. Default is TRUE.
quiet	Logical. If TRUE, suppresses progress messages.

### Details

FUNDEB (Fundo de Manutencao e Desenvolvimento da Educacao Basica e de Valorizacao dos Profissionais da Educacao) is the main funding mechanism for basic education in Brazil.

Each Excel file from STN contains ~20 data sheets named with a prefix indicating the destination (E\_ for states, M\_ for municipalities) and a suffix indicating the funding source (e.g., E\_FPE, M\_ICMS). Each sheet contains two tables: the main FUNDEB transfers and a FUNDEB adjustment table.

#### Important notes:

- Data is sourced from STN (Tesouro Nacional), not INEP.
- Files are in Excel format (XLS) — requires the readxl package.
- Column names are standardized to lowercase with underscores.
- Summary sheets (Resumo, Total, etc.) are automatically excluded.

**Value**

A tibble in tidy (long) format with columns:

**estados** State name

**uf** State code (UF)

**mes\_ano** Date (last day of the month)

**origem** Funding source (FPE, FPM, ICMS, etc.)

**destino** Transfer destination ("UF" or "Município")

**tabela** Table type ("Fundeb" or "Ajuste Fundeb")

**valor** Transfer amount in BRL (numeric)

**Data source**

<https://www.tesourotransparente.gov.br>

**See Also**

Other FUNDEB functions: [get\\_fundeb\\_enrollment\(\)](#)

**Examples**

```
## Not run:  
# get all FUNDEB distribution data for 2023  
dist_2023 <- get_fundeb_distribution(2023)  
  
# get only FPE transfers to states  
fpe_estados <- get_fundeb_distribution(2023, source = "FPE",  
                                     destination = "uf")  
  
# get data for Sao Paulo only  
sp <- get_fundeb_distribution(2023, uf = "SP")  
  
## End(Not run)
```

---

get\_fundeb\_enrollment *Get FUNDEB enrollment data*

---

### Description

Downloads and processes FUNDEB enrollment data from FNDE's OData API. These are the enrollment counts considered for FUNDEB funding calculation.

### Usage

```
get_fundeb_enrollment(  
    year,  
    uf = NULL,  
    n_max = Inf,  
    keep_file = TRUE,  
    quiet = FALSE  
)
```

### Arguments

year	The year of the data (2007-2026).
uf	Optional. A UF code (e.g., "SP", "RJ") to filter by state. The filter is applied at the API level for efficiency. Default is NULL (all states).
n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_file	Logical. If TRUE, caches the API result as a local CSV file. Default is TRUE.
quiet	Logical. If TRUE, suppresses progress messages.

### Details

Enrollment data comes from FNDE (Fundo Nacional de Desenvolvimento da Educacao) via its OData API. It includes the number of enrollments considered for FUNDEB funding, broken down by state, municipality, education type, school network, class type, and location.

#### Important notes:

- Data is sourced from FNDE, not INEP.
- Requires the `jsonlite` package.
- Results are cached locally as CSV after first download.
- Column names are standardized to lowercase with underscores.
- When `uf` is used with a cached file, filtering is done locally.

**Value**

A tibble with columns:

**ano\_censo** Census year  
**uf** State code (UF)  
**municipio** Municipality name  
**tipo\_rede\_educacao** Education network type  
**descricao\_tipo\_educacao** Education type description  
**descricao\_tipo\_ensino** Teaching type description  
**descricao\_tipo\_turma** Class type description  
**descricao\_tipo\_carga\_horaria** Class hours type description  
**descricao\_tipo\_localizacao** Location type description  
**qtd\_matricula** Number of enrollments

**Data source**

FNDE: <https://www.fnde.gov.br>

**See Also**

Other FUNDEB functions: [get\\_fundeb\\_distribution\(\)](#)

**Examples**

```
## Not run:
# get FUNDEB enrollment data for 2023
mat_2023 <- get_fundeb_enrollment(2023)

# get enrollment data for Sao Paulo only
mat_sp <- get_fundeb_enrollment(2023, uf = "SP")

# get enrollment data with limited rows
mat_sample <- get_fundeb_enrollment(2023, n_max = 1000)

## End(Not run)
```

---

get\_idd

*Get IDD (Indicador de Diferença entre os Desempenhos Observado e Esperado) data*

---

**Description**

Downloads and processes microdata from IDD, an indicator that measures the value added by an undergraduate course to student performance. It compares ENADE scores with expected performance based on students' prior achievement (ENEM scores at admission).

## Usage

```
get_idd(year, n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

## Arguments

year	The year of the indicator (2014-2019, 2021-2023). Note: there is no 2020 edition.
n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

## Details

IDD is calculated by INEP as part of the higher education quality assessment system. It complements ENADE by isolating the contribution of the course itself to student learning, controlling for student input quality.

The data includes:

- Course and institution identifiers
- IDD scores (continuous and categorical)
- Number of students considered in the calculation
- Related ENADE and ENEM metrics

### Important notes:

- IDD is published alongside ENADE results, following the same rotating cycle of course areas.
- Column names are standardized to lowercase with underscores.
- Not all courses have IDD values (minimum sample requirements apply).

## Value

A tibble with IDD data in tidy format.

## Data dictionary

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/idd>

## Examples

```
## Not run:  
# get IDD data for 2023  
idd <- get_idd(2023)  
  
# get IDD data for 2021 with limited rows  
idd_2021 <- get_idd(2021, n_max = 1000)  
  
## End(Not run)
```

---

`get_ideb`*Get IDEB (Índice de Desenvolvimento da Educação Básica) data*

---

### Description

Downloads and processes IDEB data from INEP. IDEB is the main indicator of education quality in Brazil, combining student performance (from SAEB) with grade promotion rates.

### Usage

```
get_ideb(  
  year,  
  level = c("escola", "municipio"),  
  stage = c("anos_iniciais", "anos_finais", "ensino_medio"),  
  uf = NULL,  
  quiet = FALSE  
)
```

### Arguments

<code>year</code>	The year of the IDEB (available: 2017, 2019, 2021, 2023).
<code>level</code>	The aggregation level: <ul style="list-style-type: none"><li>• "escola": School level</li><li>• "municipio": Municipality level</li></ul>
<code>stage</code>	The education stage: <ul style="list-style-type: none"><li>• "anos_iniciais": Early elementary (1st-5th grade)</li><li>• "anos_finais": Late elementary (6th-9th grade)</li><li>• "ensino_medio": High school</li></ul>
<code>uf</code>	Optional. Filter by state (UF code or abbreviation).
<code>quiet</code>	Logical. If TRUE, suppresses progress messages.

### Details

IDEB is calculated every two years since 2005 based on:

- **Learning**: Average scores in Portuguese and Mathematics from SAEB
- **Flow**: Grade promotion rate (inverse of repetition/dropout)

The index ranges from 0 to 10. Brazil's national goal is to reach 6.0 by 2022 (the level of developed countries in PISA).

**Note:** IDEB data is relatively small compared to other INEP datasets, so no `n_max` parameter is provided.

### Value

A tibble with IDEB data in tidy format.

**Data source**

Official IDEB portal: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb>

**See Also**

Other IDEB functions: [get\\_ideb\\_series\(\)](#), [list\\_ideb\\_available\(\)](#)

**Examples**

```
## Not run:
# get school-level IDEB for early elementary in 2021
ideb_escolas <- get_ideb(2021, level = "escola", stage = "anos_iniciais")

# get municipality-level IDEB for São Paulo state
ideb_sp <- get_ideb(2021, level = "municipio", stage = "anos_iniciais", uf = "SP")

# get high school IDEB for all municipalities
ideb_em <- get_ideb(2023, level = "municipio", stage = "ensino_medio")

## End(Not run)
```

---

`get_ideb_series`*Get IDEB historical series*

---

**Description**

Downloads and combines IDEB data across multiple years to create a historical series.

**Usage**

```
get_ideb_series(
  years = NULL,
  level = c("escola", "municipio"),
  stage = c("anos_iniciais", "anos_finais", "ensino_medio"),
  uf = NULL,
  quiet = FALSE
)
```

**Arguments**

<code>years</code>	Vector of years to include (default: all available).
<code>level</code>	The aggregation level.
<code>stage</code>	The education stage.
<code>uf</code>	Optional. Filter by state.
<code>quiet</code>	Logical. If TRUE, suppresses progress messages.

**Value**

A tibble with IDEB data for all requested years.

**See Also**

Other IDEB functions: [get\\_ideb\(\)](#), [list\\_ideb\\_available\(\)](#)

**Examples**

```
## Not run:  
# get IDEB history for municipalities  
ideb_hist <- get_ideb_series(  
  years = c(2017, 2019, 2021),  
  level = "municipio",  
  stage = "anos_iniciais"  
)  
  
## End(Not run)
```

---

get\_igc

*Get IGC (Indice Geral de Cursos) data*

---

**Description**

Downloads and processes IGC data from INEP. The IGC is a quality indicator for higher education institutions in Brazil, calculated as a weighted average of CPC scores across all evaluated courses plus CAPES scores for graduate programs.

**Usage**

```
get_igc(year, n_max = Inf, keep_file = TRUE, quiet = FALSE)
```

**Arguments**

year	The year of the indicator (2007-2019, 2021-2023). Note: there is no 2020 edition. Years 2004-2006 used a different indicator ("Conceito Enade").
n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_file	Logical. If TRUE, keeps the downloaded file in cache. Default is TRUE.
quiet	Logical. If TRUE, suppresses progress messages.

**Details**

IGC is calculated by INEP as part of the higher education quality assessment system (SINAES). It provides an overall quality measure for institutions, considering both undergraduate and graduate programs.

The data includes:

- Institution identifiers (code, name, organization type)
- IGC scores (continuous and categorical/faixa)
- Number of courses and students considered
- Component breakdown (undergraduate CPC average, graduate CAPES scores)

**Important notes:**

- IGC is published annually based on the last three ENADE cycles.
- There is no 2020 edition (COVID-19 suspension).
- Column names are standardized to lowercase with underscores.
- Files are in Excel format (xls/xlsx), except 2007 which is 7z.

**Value**

A tibble with IGC data in tidy format.

**Data dictionary**

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/indicadores-de-qualidade-da-educacao-s>

**See Also**

Other CPC/IGC functions: [get\\_cpc\(\)](#)

**Examples**

```
## Not run:  
# get IGC data for 2023  
igc <- get_igc(2023)  
  
# get IGC data for 2021 with limited rows  
igc_2021 <- get_igc(2021, n_max = 1000)  
  
## End(Not run)
```

---

get\_saeb

*Get SAEB (Sistema de Avaliação da Educação Básica) data*

---

**Description**

Downloads and processes microdata from SAEB, the Brazilian Basic Education Assessment System. SAEB evaluates educational quality through student performance assessments in Portuguese and Mathematics.

**Usage**

```
get_saeb(
  year,
  type = c("aluno", "escola", "diretor", "professor"),
  level = c("fundamental_medio", "educacao_infantil"),
  n_max = Inf,
  keep_zip = TRUE,
  quiet = FALSE
)
```

**Arguments**

year	The year of the assessment (2011, 2013, 2015, 2017, 2019, 2021, 2023).
type	Type of data to load. Options: <ul style="list-style-type: none"> <li>• "aluno": Student results (default)</li> <li>• "escola": School questionnaire</li> <li>• "diretor": Principal questionnaire</li> <li>• "professor": Teacher questionnaire</li> </ul>
level	For 2021 only, SAEB was split into two files: <ul style="list-style-type: none"> <li>• "fundamental_medio": Elementary and High School (default)</li> <li>• "educacao_infantil": Early Childhood Education Ignored for other years.</li> </ul>
n_max	Maximum number of rows to read. Default is Inf (all rows). Consider using a smaller value for exploration.
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

**Details**

SAEB is conducted biennially by INEP and assesses students in grades 5 and 9 of elementary school, and grade 3 of high school. The data includes:

- Student performance scores in Portuguese and Mathematics
- School infrastructure and management questionnaires
- Teacher and principal profiles

**Important notes:**

- SAEB files can be large (several hundred MB).
- Use n\_max to read a sample first for exploration.
- Column names are standardized to lowercase with underscores.
- In 2021, INEP split SAEB into two separate downloads (elementary/high school and early childhood). Use the level parameter to choose.

**Value**

A tibble with SAEB microdata in tidy format.

## Data dictionary

For detailed information about variables, see INEP's documentation: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/saeb>

## Examples

```
## Not run:  
# get student results for 2023  
saeb <- get_saeb(2023, n_max = 10000)  
  
# get school questionnaire data  
saeb_escola <- get_saeb(2023, type = "escola")  
  
# SAEB 2021: early childhood education  
saeb_infantil <- get_saeb(2021, level = "educacao_infantil", n_max = 1000)  
  
## End(Not run)
```

---

list_cache	<i>List cached files</i>
------------	--------------------------

---

## Description

Lists all files currently in the educabR cache.

## Usage

```
list_cache(dataset = NULL)
```

## Arguments

dataset            Optional. Filter by dataset name.

## Value

A tibble with information about cached files.

## See Also

Other cache functions: [clear\\_cache\(\)](#), [get\\_cache\\_dir\(\)](#), [set\\_cache\\_dir\(\)](#)

## Examples

```
## Not run:  
list_cache()  
  
## End(Not run)
```

---

list_censo_files	<i>List available Censo Escolar files</i>
------------------	---

---

**Description**

Lists the data files available in a downloaded School Census.

**Usage**

```
list_censo_files(year)
```

**Arguments**

year	The year of the census.
------	-------------------------

**Value**

A character vector of file names found.

**See Also**

Other School Census functions: [get\\_censo\\_escolar\(\)](#)

**Examples**

```
## Not run:  
list_censo_files(2023)  
  
## End(Not run)
```

---

list_censo_superior_files	<i>List available Higher Education Census files</i>
---------------------------	---

---

**Description**

Lists the data files available in a downloaded Higher Education Census. Useful for exploring the contents of the ZIP file.

**Usage**

```
list_censo_superior_files(year)
```

**Arguments**

year	The year of the census.
------	-------------------------

**Value**

A character vector of file names found.

**See Also**

Other Higher Education Census functions: [get\\_censo\\_superior\(\)](#)

**Examples**

```
## Not run:  
list_censo_superior_files(2023)  
  
## End(Not run)
```

---

`list_ideb_available`    *List available IDEB data*

---

**Description**

Lists the IDEB data files available in the INEP portal.

**Usage**

```
list_ideb_available()
```

**Value**

A tibble with available IDEB datasets.

**See Also**

Other IDEB functions: [get\\_ideb\(\)](#), [get\\_ideb\\_series\(\)](#)

**Examples**

```
list_ideb_available()
```

---

set_cache_dir	<i>Set the cache directory for educabR</i>
---------------	--

---

**Description**

Sets the directory where downloaded files will be cached. This avoids repeated downloads of the same data.

**Usage**

```
set_cache_dir(path = NULL, persistent = FALSE)
```

**Arguments**

path	A character string with the path to the cache directory. If NULL, uses a temporary directory (default).
persistent	Logical. If TRUE, the cache directory setting is saved to the user's R profile for future sessions.

**Value**

Invisibly returns the cache directory path.

**See Also**

Other cache functions: [clear\\_cache\(\)](#), [get\\_cache\\_dir\(\)](#), [list\\_cache\(\)](#)

**Examples**

```
## Not run:  
# set a custom cache directory (use tempdir() in examples)  
set_cache_dir(file.path(tempdir(), "educabR_cache"))  
  
## End(Not run)
```

# Index

- \* **CAPES functions**
    - get\_capes, 5
  - \* **CPC/IGC functions**
    - get\_cpc, 10
    - get\_igc, 25
  - \* **ENADE functions**
    - get\_enade, 11
  - \* **ENCCEJA functions**
    - get\_encceja, 13
  - \* **ENEM functions**
    - enem\_summary, 4
    - get\_enem, 14
    - get\_enem\_escola, 16
    - get\_enem\_itens, 17
  - \* **FUNDEB functions**
    - get\_fundeb\_distribution, 18
    - get\_fundeb\_enrollment, 20
  - \* **Higher Education Census functions**
    - get\_censo\_superior, 8
    - list\_censo\_superior\_files, 29
  - \* **IDD functions**
    - get\_idd, 21
  - \* **IDEB functions**
    - get\_ideb, 23
    - get\_ideb\_series, 24
    - list\_ideb\_available, 30
  - \* **SAEB functions**
    - get\_saeb, 26
  - \* **School Census functions**
    - get\_censo\_escolar, 7
    - list\_censo\_files, 29
  - \* **cache functions**
    - clear\_cache, 3
    - get\_cache\_dir, 5
    - list\_cache, 28
    - set\_cache\_dir, 31
- available\_years, 2
- clear\_cache, 3, 5, 28, 31
- enem\_summary, 4, 15–17
- get\_cache\_dir, 3, 5, 28, 31
- get\_capes, 5
- get\_censo\_escolar, 7, 29
- get\_censo\_superior, 8, 30
- get\_cpc, 10, 26
- get\_enade, 11
- get\_encceja, 13
- get\_enem, 4, 14, 16, 17
- get\_enem(), 4
- get\_enem\_escola, 4, 15, 16, 17
- get\_enem\_itens, 4, 15, 16, 17
- get\_fundeb\_distribution, 18, 21
- get\_fundeb\_enrollment, 19, 20
- get\_idd, 21
- get\_ideb, 23, 25, 30
- get\_ideb\_series, 24, 24, 30
- get\_igc, 11, 25
- get\_saeb, 26
- list\_cache, 3, 5, 28, 31
- list\_censo\_files, 8, 29
- list\_censo\_superior\_files, 9, 29
- list\_ideb\_available, 24, 25, 30
- set\_cache\_dir, 3, 5, 28, 31