Package: bvq (via r-universe)

October 29, 2024

Title Barcelona Vocabulary Questionnaire Database and Helper Functions

Version 0.4.0

Maintainer Gonzalo Garcia-Castro <gongarciacastro@gmail.com>

Description Download, clean, and process the Barcelona Vocabulary
Questionnaire (BVQ) data. BVQ is a vocabulary inventory
developed for assesing the vocabulary of Catalan-Spanish
bilinguals infants from the Metropolitan Area of Barcelona
(Spain). This package includes functions to download the data
from formr servers, and return the processed data in multiple
formats.

License MIT + file LICENSE

URL https://gongcastro.github.io/bvq/,

https://github.com/gongcastro/bvq/

BugReports https://github.com/gongcastro/bvq/issues

Depends R (>= 3.5.0),

Imports cli (>= 3.6.1), dplyr (>= 1.1.0), formr (>= 0.9.1),
googlesheets4 (>= 1.0.0), janitor (>= 2.2.0), lifecycle (>=
1.0.3), lubridate (>= 1.8.0), magrittr (>= 2.0.3), purrr (>=
1.0.1), readxl (>= 1.4.3), rlang (>= 1.1.1), tibble (>= 3.2.1),
tidyr (>= 1.2.0)

Suggests knitr, rmarkdown, roxygen2, testthat (>= 3.0.0), ggplot2 (>=
3.4.2), withr

VignetteBuilder knitr

Remotes github::rubenarslan/formr

Config/testthat/edition 3

Encoding UTF-8

Language en-US

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Repository https://gongcastro.r-universe.dev

RemoteUrl https://github.com/gongcastro/bvq

RemoteRef HEAD

RemoteSha 5b7fb8d6571d5039908f825a14c76792d3a54475

Contents

---

bvq_connect   *Authenticate in Google and formr*

Description

This function tries to log in to the formr API with the user-provided password (argument `password`) or retrieving it from the global environment (`FORMR_PWD` in .Renviron)

Usage

```
bvq_connect(google_email = NULL, password = NULL)
```

Arguments

google_email   E-mail used in Google Drive account. If `NULL` (default), it is assumed to be the same as `formr_email`.

password   Character string with the password to formr (`NULL` by default).

Value

Logical. `TRUE` if Google and formr authentication was successful, `FALSE` if authentication of any of the two failed.

Examples

```
## Not run:
bvq_connect()

## End(Not run)
```

---

bvq_items   *Get items included in a questionnaire*

Description

This function retrieves information about the items in a particular section in the BVQ questionnaire.
This includes item names, item types, text, choices, settings, and other metadata.

Usage

```
bvq_items(section, version = "bvq-1.0.0")
```

Arguments

section   Name of the version of the questionnaire for which the items of a section will
                 be retrieved. Check the output of names(get_bvq_runs()) to see the available
                 versions of the questionnaire.

Value

A list of length 3, which includes:

- survey: A tibble::tibble containing the items included in the questionnaire and several prop-
  erties. Each row corresponds to a single item, and each column corresponds to a particular
  property:
  - type: a character string indicating the type of the item (see formr documentation).
  - name: a character string indicating the name of the item, as it appears in the output of
    bvq_responses().
  - label: a character string indicating the text shown to participants when filling out the
    questionnaire.
  - optional: a logical value indicating whether providing an answer to the item is mandatory
    for participants.
  - class: a character string indicating the CSS class with of the item.
  - showif: a character string indicating R code that determines under what conditions the
    item is shown to participants.
  - value: default value of the item.
  - block_order: character string (a latter) indicating the order in which the block that the
    item belongs to appears in the survey.

   – item_order: integer indicating the order in which the item appears within the block it
      belongs to belongs.
• choices: A tibble::tibble containing the choices given to participants for some items. Each
   row corresponds to a choice, and each column corresponds to a particular choice property:
   – list_name': character string indicating the name of the name of the choice list (which may
      repeat across different items).
   – name: character string indicating the name that a particular choice will be assigned in the
      code.
   – label: character string indicating the text that will be show to participants for a particular
      choice.
• settings: A tibble::tibble containing the settings for the survey. Each row corresponds to one
   setting, and each column indicates the setting names and values:
   – item: name of the setting.
   – value: value of the setting.

Author(s)

Gonzalo Garcia-Castro

Examples

```
## Not run:
bvq_items("bvq_06_words_catalan", version = "bvq-1.0.0")

## End(Not run)
```

---

bvq_logs   *Generate participant information and progress for each response*

Description

This function generates a data frame that contains participant-level information. Each row is a
given participant's response and each column is a variable. The same participant will always be
identified with the same id. The variable time indexes how many times a participant has been sent
the questionnaire, independently of whether a response was obtained from them later.

Usage

```
bvq_logs(
  participants = bvq_participants(),
  responses = bvq_responses(participants),
  bilingual_threshold = 0.8,
  other_threshold = 0.1
)
```

Arguments

participants   Participants data frame, as generated by `bvq_participants()`.

responses   Responses data frame, as generated by `bvq_responses()`.

bilingual_threshold

Numeric scalar ranging from 0 to 1 indicating the minimum degree of exposure to Catalan or Spanish to consider a participant as *Monolingual*.

other_threshold

Numeric scalar ranging from 0 to 1 indicating the minimum degree of exposure to languages other than Catalan and Spanish to consider a participant as *Other*.

Value

A data frame (actually, a tibble::tibble with participant-level information. Each row corresponds to a questionnaire response and each column represents a variable. The output includes the following variables:

- child_id: a character string with five digits indicating a participant's identifier in the database from the Laboratori de Recerca en Infància at Universitat Pompeu Fabra. This value is always the same for each participant, so that different responses from the same participant share the same `id`.

- response_id: a character string identifying a single response to the questionnaire. This value is always unique for each response to the questionnaire, even for responses from the same participant.

- time: a numeric value indicating how many times a given participant has been sent the questionnaire, regardless of whether they completed it or not.

- study: a character string indicating the study in which the participant was invited to fill in the questionnaire. Frequently, participants that filled in the questionnaire came to the lab to participant in a study, and were then invited to fill in the questionnaire later. This value indicates what study each participant was tested in before being sent the questionnaire.

- version: a character string indicating what version of the questionnaire a given participant filled in. Different versions may contain a different subset of items, and the administration instructions might vary slightly (see formr questionnaire templates in the [GitHub repository(https://github.com/gongcastro/multilex)). Also, different versions were designed, implemented, and administered at different time points (e.g., before/during/after the COVID-related lockdown).

- version_list: a character string indicating the specific list of items a participant was assigned to. Only applies in the case of short versions of BVQ, such as bvq-short, bvq-long, bvq-lockdown, or bvq-1.0.0, where the list of items was partitioned into several versions.#' * date_sent: a date value (see lubridate package) in yyyy/mm/dd format indicating the date in which the questionnaire was sent to participants.

- days_from_sent: a numeric value indicating the number of days elapsed since participants were sent the questionnaire (as indicated by `date_sent`) and completed the questionnaire.

- date_birth: a date value (see lubridate package) in yyyy/mm/dd format indicating participants birth date.

- age: a numeric value indicating the number of months elapsed since participants' birth date until they filled in the last item of their questionnaire response.

- age_today: a numeric value indicating the number of months elapsed since participants' birth date until the present day, as indicated by lubridate::now.

- months_from_last_response: a numeric value indicating the number of months elapsed since participants' last questionnaire response (as indicated by time_stamp) until the present day, as indicated by lubridate::now.

- edu_parent1: a character string indicating the educational attainment of one of the parents/caregivers.

- edu_parent2: a character string indicating the educational attainment of the other parent/caregiver, if any.

- dominance: a character string indicating the language of highest exposure ("Catalan" or "Spanish"), as reported by parents. If exposure is identical for both language, "Catalan" is assigned.

- lp: a character string indicating participants' language profile, classified using parental reports of language exposure (see doe_spanish, doe_catalan, and doe_others), and the thresholds passed in the bilingual_threshold and other_threshold.

- doe_spanish: a numeric value ranging from 0 to 1 indicating participants' daily exposure to Spanish, as estimated by parents/caregivers. This value aggregates participants' exposure to any variant of Spanish (e.g., European and American Spanish).

- doe_catalan: a numeric value ranging from 0 to 1 indicating participants' daily exposure to Catalan, as estimated by parents/caregivers. This value aggregates participants' exposure to any variant of Catalan (e.g., Catalan from Mallorca or Barcelona).

- doe_others: a numeric value ranging from 0 to 1 indicating participants' daily exposure to languages other than Spanish or Catalan, as estimated by parents/caregivers, aggregating participants' exposure to all those other languages (e.g., Norwegian, Arab, Swahili).

- completed: a logical value that returns TRUE if progress is 1, and FALSE otherwise.

Author(s)

Gonzalo Garcia-Castro

Examples

```
## Not run:
responses <- bvq_responses()

logs <- bvq_logs(responses = responses)

## End(Not run)
```

---

bvq_norms	*Generate item-level norms for age, sex, language profile and item dominance*

Description

This function generates a data frame with the estimated proportion of children that understand and/or produce some items for a selected age range and participant profiles. Estimated proportions and corresponding standard errors and confidence intervals are computed adjusting for zero- and one-inflation (see function `prop_adj()`).

Usage

```
bvq_norms(
  participants = bvq_participants(),
  responses = bvq_responses(participants),
  ...,
  te = NULL,
  item = NULL,
  age = c(0, Inf)
)
```

Arguments

participants	Participants data frame, as generated by `bvq_participants()`
responses	Responses data frame, as generated by `bvq_responses()`.
...	<dynamic-dots>. Unquoted name of the variable(s) to group data into. Norms will be calculated by aggregating responses within the groups that result from the combination of crossing of the variables provided in .... These variables can refer to item properties (see pool, e.g., `semantic_category`) or to participant properties (see `bvq_logs()`, e.g., `lp`).
te	Translation equivalent for which the norms should be computed.
	• If NULL (default), norms are computed exclusively for the items indicated in `item`.
	• If TRUE, norms are computed for both the item indicated in the `item` argument, and for its translation.
	• If FALSE, norms are computed exclusively for the items indicated in `item` (same as NULL)
	• If numeric vector, norms are computed for all items corresponding to the translation equivalents indicated in this argument.
item	Character string indicating the item to compute norms for. If left `NULL` (default) norms will be computed for all items. You can check the available items in the pool data set running `data("pool")`.
age	Numeric vector of length two (*min-max*) indicating the age range of participants to compute norms for.

Value

A data frame (actually, a tibble::tibble with the proportion of participants in the sample that understand or produce the items indicated in `item`, or the translation equivalents indicated in `te`. The output contains the following variables:

- te: an integer identifying the translation equivalent (a.k.a., pair of cross-language synonyms, doublets) the item belongs to.

- item: character string indicating the item identifier (e.g., spa_mesa). This value is unique for each item. Responses to the same item from different participants are linked by the same item value.

- language: a character string indicating the language the item response belongs to: "Catalan" if item in Catalan), "Spanish" if item in Spanish.

- age: an numeric vector of length 1 or 2 indicating the age range of participants (in months) for which the estimates should be computed. If a non-integer is provided (e.g., 15.36, it is rounded downwards using floor().)

- type: a character string indicating the vocabulary type computed: "understands" if option 'Understands' was selected, and "produces" if option 'Understands & Says' was selected.

- item_dominance: a character string that takes the value "L1" if the item belongs to participants' language of most exposure, and L2 if the item belongs to participants' language of least exposure.

- label: a character string indicating the text presented to participants in the questionnaire (replacing the item identifier).

- .sum: a positive integer indicating the number of positive responses: responses is 2 (Understands) or 3 (Understands & Says) for type = "understands", and 3 (Understands & Says) if type = "produces".

- .n: a positive integer indicating the total number number of responses (useful for computing proportions).

- .prop: a numeric value ranging from 0 to 1 (both included) indicating the estimated proportion of participants that provided a positive response, adjusted following Gelman et al.'s method to account for zero- and one-inflation (see function prop_adj). Additionally, any variables specified in the .by argument are preserved as grouping variables.

Author(s)

Gonzalo Garcia-Castro

Examples

```
## Not run:
responses <- bvq_responses()

bvq_norms(
  participants = participants,
  responses = responses,
  item = "cat_casa",
  age = c(22, 22),
  lp
)

my_items <- c("cat_gos", "cat_gat")

bvq_norms(
```

```
    participants = participants,
    responses = responses,
    item = my_items,
    te = TRUE,
    age = c(15, 16)
)

## End(Not run)
```

---

bvq_participants   *Retrieve and update local and/or remote data from formr*

Description

This function generates a data frame with the information of all participants that have participated or are candidates to participate in any of the versions of BVQ.

Usage

```
bvq_participants(...)
```

Arguments

...   Unused.

Value

A data frame (actually, a tibble::tibble) with all participants that have participated or are candidates to participate in any of the versions of BVQ Each row corresponds to a questionnaire response and each column represents a variable. The output includes the following variables:

- child_id: a character string with five digits indicating a participant's identifier in the database from the Laboratori de Recerca en Infància at Universitat Pompeu Fabra. This value is always the same for each participant, so that different responses from the same participant share the same id.

- response_id: a character string identifying a single response to the questionnaire. This value is always unique for each response to the questionnaire, even for responses from the same participant.

- time: a numeric value indicating how many times a given participant has been sent the questionnaire, regardless of whether they completed it or not.

- date_birth: a date value in yyyy/mm/dd format indicating participants birth date.

- age_now: a numeric value indicating the number of months elapsed since participants' birth date until the present day, as indicated by lubridate::now().

- version: a character string indicating what version of the questionnaire a given participant filled in. Different versions may contain a different subset of items, and the administration instructions might vary slightly (see formr questionnaire templates in the GitHub repository. Also, different versions were designed, implemented, and administrated at different time points (e.g., before/during/after the COVID-related lockdown).

- version_list: a character string indicating the specific list of items a participant was assigned to. Only applies in the case of short versions of BVQ, such as bvq-short, bvq-long, bvq-lockdown, or bvq-1.0.0, where the list of items was partitioned into several versions.

- date_test: a date value (see lubridate package) in yyyy/mm/dd format indicating the date in which the participant was tested in the associated study, if any.

- date_sent: a date value (see lubridate package) in yyyy/mm/dd format indicating the date in which the participant was sent the questionnaire.

- call: a character string indicating the status of the participant's response:
  - "successful": participant completed the questionnaire)
  - "sent": participant has been sent the email but has not completed it yet)
  - "pending": participant is still to be sent the questionnaire.
  - "reminded": a week has elapsed since the participant was sent the questionnaire, and has been already reminded of it.
  - "stop": participant has not completed the questionnaire after two weeks since they were sent the questionnaire.

Author(s)

Gonzalo Garcia-Castro

Examples

```
## Not run:
bvq_participants()

## End(Not run)
```

---

bvq_responses	*Retrieve and update local and/or remote data from formr*

Description

This function generates a data frame with participant's responses to each item, along with some session-specific metadata. It takes participants (the output of bvq_participants()) and runs (a character vector that can take zero, one, or multiple of the following values: "formr2", "formr-short", "formr-lockdown") as arguments.

Usage

```
bvq_responses(participants = bvq_participants())
```

Arguments

participants   Participants data frame, as generated by `bvq_participants()`. If NULL (default), `bvq_participants()` is run.

Value

A data frame (actually, a tibble::tibble containing participant's responses to each item, along with some session-specific metadata. The output includes the following variables:

- child_id: a character string with five digits indicating a participant's identifier in the database from the Laboratori de Recerca en Infància at Universitat Pompeu Fabra. This value is always the same for each participant, so that different responses from the same participant share the same child_id.

- response_id: a character string identifying a single response to the questionnaire. This value is always unique for each response to the questionnaire, even for responses from the same participant.

- time: a numeric value indicating how many times a given participant has been sent the questionnaire, regardless of whether they completed it or not.

- version: a character string indicating what version of the questionnaire a given participant filled in. Different versions may contain a different subset of items, and the administration instructions might vary slightly (see formr questionnaire templates in the GitHub repository). Also, different versions were designed, implemented, and administrated at different time points (e.g., before/during/after the COVID-related lockdown).

- version_list: a character string indicating the specific list of items a participant was assigned to. Only applies in the case of short versions of BVQ, such as bvq-short, bvq-long, bvq-lockdown, or bvq-1.0.0, where the list of items was partitioned into several versions.#' * item: character string indicating the item identifier (e.g., spa_mesa). This value is unique for each item. Responses to the same item from different participants are linked by the same item value.

- response: integer indicating the participant's response to a give item: 1 if "No" (the participant does not understand or produce the word), 2 if "Understands" (the participants understands the word), or 3 if "Understands and Says" (the participant understands and produces the item).

- date_birth: lubridate::Date indicating participants birth date.

- date_started: lubridate::Date indicating when participants logged to the questionnaire for the first time.

- date_finished: lubridate::Date indicating when participants logged to the questionnaire for the last time.

- sex: a character string indicating participants' biological sex, as reported by the parents.

- doe_spanish: a numeric value ranging from 0 to 1 indicating participants' daily exposure to Spanish, as estimated by parents/caregivers This value aggregates participants' exposure to any variant of Spanish (e.g., European and American Spanish).

- doe_catalan: a numeric value ranging from 0 to 1 indicating participants' daily exposure to Catalan, as estimated by parents/caregivers This value aggregates participants' exposure to any variant of Catalan (e.g., Catalan from Majorca or Barcelona).

• doe_others: a numeric value ranging from 0 to 1 indicating participants' daily exposure to languages other than Spanish or Catalan, as estimated by parents/caretakers, aggregating participants' exposure to all those other languages (e.g., Norwegian, Arab, Swahili).

• edu_parent1: a character string indicating the educational attainment of one of the parents/caretakers.

• edu_parent2: a character string indicating the educational attainment of the other parent/caretaker, if any.

Author(s)

Gonzalo Garcia-Castro

Examples

```
## Not run:
bvq_responses()

## End(Not run)
```

---

bvq_vocabulary   *Generate participant information and progress for each response*

Description

This function generates a data frame with the vocabulary of each participant (keeping longitudinal data from the same participant in different rows). Comprehensive and productive vocabulary sizes are computed as raw counts (*_count) and as proportions *_prop.

Usage

```
bvq_vocabulary(
  participants = bvq_participants(),
  responses = bvq_responses(participants),
  ...,
  .scale = "prop"
)
```

Arguments

participants   Participants data frame, as generated by [bvq_participants()](#).

responses   Responses data frame, as generated by [bvq_responses()](#).

...   [<dynamic-dots>](#) Unquoted name of the variable(s) to group data into. Vocabulary metrics will be calculated by aggregating responses within the groups that result from the combination of crossing of the variables provided in .... These variables can refer to item properties (see [pool](#), e.g., semantic_category) or to participant properties (see [bvq_logs()](#), e.g., lp).

.scale   A character vector that takes the value "count" and/or "prop". If "prop" (default), vocabulary metrics are calculated as proportions. If "count", vocabulary metrics are reported as counts (number of words).

Value

A dataset (actually, a tibble::tibble with each participant's comprehensive and/or vocabulary size in each language. This data frame contains the following variables:

- child_id: a character string with five digits indicating a participant's identifier in the database from the Laboratori de Recerca en Infància at Universitat Pompeu Fabra. This value is always the same for each participant, so that different responses from the same participant share the same child_id.

- response_id: a character string identifying a single response to the questionnaire. This value is always unique for each response to the questionnaire, even for responses from the same participant.

- age: a numeric value indicating the number of months elapsed since participants' birth date until they filled in the last item of their questionnaire response.

- type: a character string indicating the vocabulary type computed: "understands" if option "Understands" was selected, and "produces" if option "Understands & Says" was selected.

- total_count: integer indicating the number of items selected as "Understands" or "Understands and Says" in both languages.

- l1_count: positive integer indicating the number of items selected as "Understands" or "Understands and Says" in the dominant language (L1).

- l2_count: positive integer indicating the number of items selected as "Understands" or "Understands and Says" in the non-dominant language (L2).

- concept_count: positive integer indicating the number of translation equivalents (a.k.a. cross-language synonyms or doublets) in which "at list one of the items was selected as "Understands" or "Understands and Says". This is a measure of the number of lexicalised concepts.

- te_count: positive integer indicating the number of translation equivalents (out of the total number of items the participant answered to) in which at both items was selected as "Understands" or "Understands and Says". This is a measure of the number of lexicalised concepts.

- total_prop: numeric value ranging from 0 to 1 (both included) indicating the proportion of items selected as "Understands" or "Understands and Says" in both languages.

- l1_prop: numeric value ranging from 0 to 1 (both included) indicating the proportion of of items selected as "Understands" or "Understands and Says" in the dominant language (L1).

- l2_prop: numeric value ranging from 0 to 1 (both included) indicating the proportion of of items selected as "Understands" or "Understands and Says" in the non-dominant language (L2).

- concept_prop: numeric value ranging from 0 to 1 (both included) indicating the proportion of of translation equivalents (a.k.a. cross-language synonyms or doublets) in which at least one of the items was selected as "Understands" or "Understands and Says". This is a measure of the number of lexicalised concepts.

- te_prop: numeric value ranging from 0 to 1 (both included) indicating the proportion of of translation equivalents (aka. cross-language synonyms or doublets) in which both items were selected as "Understands" or "Understands and Says". This is a measure of the number of lexicalised concepts. The specific subset of columns returned by bvq_vocabulary() depends on the elements of ... and .scale.

- contents: list containing the items marked as acquired.

Author(s)

Gonzalo Garcia-Castro

---

get_doe   *Summarise language profile*

Description

Summarise language profile

Usage

```
get_doe(...)
```

Arguments

...   Columns with the degree of exposures to be summed up for (all others will be considered as doe_others).

Value

A numeric vector with the row-wise sums of the columns specified in . . . .

Author(s)

Gonzalo Garcia-Castro

Examples

```
library(dplyr)
x <- data.frame(
  doe_cat_1 = seq(0, 1, 0.1),
  doe_cat_2 = c(0, rep(c(0.1, 0), each = 5)),
  doe_spa_1 = c(0, rep(c(0.1, 0), each = 5)),
  doe_spa_2 = c(1, 0.7, 0.6, 0.5, 0.3, 0.1, 0.4, 0.3, 0.2, 0.1, 0)
)

y <- mutate(x,
  doe_other = 1 - get_doe(matches("cat|spa")),
  doe_cat = get_doe(doe_cat_1, doe_cat_2),
  doe_spa = get_doe(matches("spa"))
)

(y)
```

---

get_longitudinal   *Deal with repeated measures*

Description

Deal with repeated measures

Usage

```
get_longitudinal(x, longitudinal = "all")
```

Arguments

x   A data frame containing a column for participants (each participant gets a unique ID), and a column for times (a numeric value indicating how many times each participant appears in the data frame counting this one). One participant may appear several times in the data frame, with each time with a unique value of `time`.

longitudinal   A character string indicating what subset of the participants should be returned:

- `"all"` (default) returns all participants.
- `"no"` remove all participants with more than one response.
- `"only"` returns only participants with more than one response in the dataset (i.e., longitudinal participants).
- `"first"` returns the first response of each participant (participants with only one appearance are included).
- `"last"` returns the last response from each participant (participants with only one response are included).

Value

A subset of the data frame `x` with only the selected cases, according to `longitudinal`.

Author(s)

Gonzalo Garcia-Castro

Examples

```
child_id <- c(1, 1, 1, 2, 2, 3, 4, 4, 4, 4, 5, 6, 7, 7, 8, 9, 10, 10)
sums <- rle(sort(child_id))[["lengths"]]
dat <- data.frame(child_id, time = unlist(sapply(sums, function(x) seq(1, x))))

(dat)

get_longitudinal(dat, "first")
get_longitudinal(dat, "only")
```

---

launch_app   *Launch bvq Shiny App in a browser*

Description

Launch bvq Shiny App in a browser

Usage

```
launch_app()
```

Details

The BVQ Shiny App provides a visual interface to the bvq R package to explore the database. Its GitHub repository contains the data, documentation, and R scripts needed to run the BVQ Shiny app.

Source

https://github.com/gongcastro/bvq-app

---

pool   *Pool of words*

Description

A dataset containing candidate words to be included in the questionnaires with some lexical properties. Transcriptions were (a) generated manually, (b) retrieved from Wiktionary. All transcriptions have been manually double-checked and fixed if necessary.

Usage

```
pool
```

Format

A data frame with 1601 rows and 20 variables:

- item: item label, as indicated in the formr survey spreadsheets, items are unique within and across questionnaires.
- language: language the item belongs to.
- te: index associated to translation equivalents across languages.
- label: item label, as presented to participants in the front-end of the questionnaire, some labels are not unique within or across questionnaires.
- xsampa: phonological transcription in X-SAMPA format.

- n_lemmas: an integer indicating the number of different lemmas showed in the item label to participants. for instance, the Spanish item `"spa_hierba"` was shown to in the questionnaire as `"hierba / césped"`. Lemma with similar roots were considered as one, such as the Spanish item `"spa_tonto"`, presented as `"tonto / tonta"` in the questionnaire.

- is_multiword: an logical indicating whether the item included a multi-word phrase as presented in the questionnaire. For instance the Spanish item `"spa_cepillodientes"` was shown as `"cepillo de dientes"` in the questionnaire, which includes three words.

- subtlex_lemma: word label, as included in the corresponding version. of SUBTLEX.

- wordbank_lemma: word label, as indexed in Wordbank.

- childes_lemma: word label, as it appears in the CHILDES English corpora (based on `wordbank_lemma`).

- semantic_category: semantic/functional category the items belongs to.

- class: Functional category (verb, nouns, adjective, etc.).

- version: what short version of the questionnaire does this item appear on?

- include: should this item be included in analyses?

---

prop_adj	*Proportion, adjusted for zero- and one-inflation*

Description

Proportion, adjusted for zero- and one-inflation

Usage

```
prop_adj(x, n)
```

Arguments

x	Number of successes
n	Number of tries

Details

It is very common that a large proportion of the participants know or do not know some word. Vocabulary sizes and word prevalence norms in package are calculated using an estimate that adjusts for zero- and one-inflation so that, at the population level such estimates are more likely to be accurate.

Value

A numeric scalar.

Examples

```
prop_adj(4, 60)
```

---

track_progress   *Track a participant's response progress.*

Description

This function prints some informative messages about a participants progress through the BVQ, and returns a vector of logical values indicating the surveys that the participant has completed.

Usage

```
track_progress(response_id, participants = NULL, ...)
```

Arguments

participants	Participants data frame, as generated by bvq_participants().
...	Arguments passed to download_surveys.
respose_id	a character string identifying a single response to the questionnaire. This value is always unique for each response to the questionnaire, even for responses from the same participant.

Value

A logical vector indicating the surveys that the participant has completed.

Author(s)

Gonzalo Garcia-Castro

Examples

```
## Not run:
track_progress("1911", participants, verbose = FALSE)

## End(Not run)
```

Index