

---

## What happens in the pretest: An empirical investigation of pretesting effects in incidental L2 vocabulary learning.

*A Data Management Plan created using DMPonline.be*

**Creators:** Eva Puimège, n.n. n.n.

**Affiliation:** KU Leuven (KUL)

**Funder:** Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

**Template:** FWO DMP (Flemish Standard DMP)

**Principal Investigator:** Eva Puimège

**Data Manager:** Eva Puimège, n.n. n.n.

**Project Administrator:** Eva Puimège, n.n. n.n.

**Grant number / URL:** 1213023N

**ID:** 199285

**Start date:** 01-10-2022

**End date:** 30-09-2025

**Last modified:** 03-05-2023

# What happens in the pretest: An empirical investigation of pretesting effects in incidental L2 vocabulary learning.

## Application DMP

---

### Questionnaire

**Describe the datatypes (surveys, sequences, manuscripts, objects ... ) the research will collect and/or generate and /or (re)use. (use up to 700 characters)**

The project will generate new data, including personal data.

1. Vocabulary instruments: digital vocabulary tests made in the Gorilla experiment builder (gorilla.sc), including the Vocabulary Levels Test (Schmitt et al., 2001), and the pre- and posttests of the three experiments (data type = online survey/experiment, format = html)
2. Vocabulary test results: spreadsheets containing the results of the pre- and posttests and the Vocabulary Levels Test, generated by Gorilla (data type = experimental/observational, format = .csv, .xls)
3. Eye-tracking experiments: Experiment Builder files, deployed experiment, interest area files, excel containing the data source, other small files generated as part of the experiment (data type = software, format = .csv, .xls, .txt, .ebd, .exe)
4. Eye-tracking data: raw eye-tracking data, data viewer projects, and spreadsheets containing cleaned data (data type = experimental, format = .csv, .xls, .edf, .evs, .xls)
5. Questionnaires: demographic surveys and reading task questionnaires made in the Gorilla experiment builder, including rating scales and multiple-choice questions (data type = survey/questionnaire, format = html)
6. Consent form: consent form made in Gorilla experiment builder (data type = survey/questionnaire, format = html)
7. Questionnaire responses: spreadsheets containing responses to background questionnaires (demographic data), and questionnaire responses related to the reading task and the experimental conditions, generated by Gorilla (data type = observational, format = .csv, .xls)
8. Statistical data: R codes, figures, spreadsheets containing aggregated data, and results in text and spreadsheets (data type = compiled/aggregated data, format = .csv, .txt, .xls, .r, .tif)
9. Output for publication: article drafts and manuscripts, conference abstracts, cover letters, responses to reviews (data type = compiled/aggregated data, format = .docx, .pdf)
10. Accompanying information: Documentation and metadata about study procedures and data storage (format = .r, .docx)

**Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)**

1. Responsible person: Prof. Elke Peters, elke.peters@kuleuven.be
2. Storage capacity/repository:
  - during the research, all data will be stored on the university's central servers via OneDrive. OneDrive is a safe university-wide platform that has automatic daily back-up procedures. As no big data sets will be generated as part of the project (< 1 TB), OneDrive has enough storage capacity for all data generated as part of the project.
  - after the research, personal data will likewise be stored on the university's central servers (with automatic back-up procedures) for 10 years, after which the data will be deleted.
  - materials generated for the project (tests, eye-tracking experiments, etc.) will be made publicly available via the Open Science Framework (OSF). Results of the study (data generated for individual participants, including test results and eye-tracking results) will only be published in summarized form, full data sets or raw data will not be shared beyond the data sharing contract with Tohoku University (Japan).

**What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)**

NA

**Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)**

Personal data will be collected from human participants: these data include the following categories:

- demographic information such as age, gender, and level of education
- bank details, names, and addresses (for the payment of participants)
- email addresses (for the recruitment of participants)

In order to protect the participants' privacy, data will be pseudonymized. Each participant will be assigned a randomized ID code that will be used in all data processing and analysis. Information that can identify a participant (e.g., name, email address) will be stored in a single file (.xls) along with the corresponding ID code. After data collection, this file will only be consulted if a participant wants to exercise their right of access to the data. Personal data such as age and gender will only be shared in aggregated form. Personal data that can identify individual participants (e.g., name, email address) or bank details will only be shared with the financial antenna for payment purposes.

Identification data (name, email address) and data collected for the payment of participants (name, address, account number) will not be included in any of the online surveys or experiments (e.g., in Gorilla) but will be kept separate from the experimental data.

Finally, for Study 2 of the project, data will be collected from participants in Japan and Belgium. Data collection will therefore involve a collaboration with between the KU Leuven researchers and researchers at Tohoku University in Japan. A data transfer agreement will be drawn up stating that only pseudonymized data will be transferred between universities.

**Which other issues related to the data management are relevant to mention? (use up to 700 characters)**

NA

## **What happens in the pretest: An empirical investigation of pretesting effects in incidental L2 vocabulary learning.**

### **DPIA**

---

#### **DPIA**

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

**What happens in the pretest: An empirical investigation of pretesting effects in incidental L2 vocabulary learning.**

**GDPR**

---

**GDPR**

Have you registered personal data processing activities for this project?

- Not applicable

# What happens in the pretest: An empirical investigation of pretesting effects in incidental L2 vocabulary learning.

## FWO DMP (Flemish Standard DMP)

### 1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format (MB/GB/TB)	Physical volume
Vocabulary tests	Digital vocabulary tests made in the Gorilla experiment builder (gorilla.sc), including the Vocabulary Levels Test (Schmitt et al., 2001), and the pre- and posttests of the three experiments	Generate new data	Digital	Other: online survey/experiment	.html	<1GB
Vocabulary test results	Spreadsheets containing the results of the pre- and posttests and the Vocabulary Levels Test, generated by Gorilla	Generate new data	Digital	Experimental/observational	.csv, .xls	<10GB
Eye-tracking experiments	Experiment Builder files, deployed experiment, interest area files, excel containing the data source, other small files generated as part of the experiment	Generate new data	Digital	Software	.csv, .xls, .txt, .ebd, .exe	< 10GB
Eye-tracking data	Raw eye-tracking data, data viewer projects, and spreadsheets containing cleaned data	Generate new data	Digital	Experimental	.csv, .xls, .edf, .evs, .xls	< 100GB
Questionnaires	Demographic surveys, consent forms and reading task questionnaires made in the Gorilla experiment builder, including rating scales and multiple-choice questions	Generate new data	Digital	Survey/questionnaire	.html	< 1GB
Questionnaire responses	Spreadsheets containing responses to background questionnaires (demographic data), and questionnaire responses related to the reading task and the experimental conditions, generated by Gorilla	Generate new data	Digital	Observational	.csv, .xls	< 10GB
Statistical data	R codes, figures, spreadsheets containing aggregated data, and results in text and spreadsheets	Generate new data	Digital	Compiled/aggregated data	.csv, .txt, .xls, .r, .tif	< 10GB
Accompanying information	Documentation and metadata about study procedures and data storage	Generate new data	Digital	Other	.r, .docx	< 1GB

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

NA

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

Ethical approval was obtained from PRET for the first study of the project ( [G-2022-5937](#) ). The KU Leuven compliance monitoring form number is E-2022-3273. This form will be updated for studies 2 and 3 of the project.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

Personal data will be collected from human participants: these data include the following categories:

- demographic information such as age, gender, and level of education
- bank details, names, and addresses (for the payment of participants)
- email addresses (for the recruitment of participants)

In order to protect the participants' privacy, data will be pseudonymized. Each participant will be assigned a randomized ID code that will be used in all data processing and analysis. Information that can identify a participant (e.g., name, email address) will be stored in a single file (.xls) along with the corresponding ID code. After data collection, this file will only be consulted if a participant wants to exercise their right of access to the data. Personal data such as age and gender will only be shared in aggregated form. Personal data that can identify individual participants (e.g., name, email address) or bank details will only be shared with the financial antenna for payment purposes.

Identification data (name, email address) and data collected for the payment of participants (name, address, account number) will not be included in any of the online surveys or experiments (e.g., in Gorilla) but will be kept separate from the experimental data.

Finally, for Study 2 of the project, data will be collected from participants in Japan and Belgium. Data collection will therefore involve a collaboration with between the KU Leuven researchers and researchers at Tohoku University in Japan. A data transfer agreement will be drawn up stating that only pseudonymized data will be transferred between universities.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, ...)? If so, please comment per dataset or data type where appropriate.

- No

NA

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

For Study 2 of the project, data will be collected from participants in Japan and Belgium. Data collection will therefore involve a collaboration with between the KU Leuven researchers and researchers at Tohoku University in Japan. A data transfer agreement will be drawn up stating that only pseudonymized data will be transferred between universities.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

NA

## 2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

All data will be stored in the project folder. The name of the folder contains the project ID. Subfolders will be made for different types of data (text data and eye-tracking data, statistical data, etc.). The project folder will contain a codebook containing details about the study design, methodology, data coding and analysis. The folder also contains the information letter (Word document), the informed consent form (Word document), debriefing letter (Word document), and a description of the informed consent procedure (Word document). Finally, the folder will contain a description of the pseudonymization procedure in a Word document.

Eyelink projects and raw eye-tracking data will be stored along with a txt file that describes what the data represent and how they were generated. The subfolder also contains the lab book, which describes lab setup, eyetracker settings, and host and display PC settings. The lab book will also describe the data collection procedure, including participant instructions. Finally, the lab book will contain notes taken during data collection, with time stamps and participant IDs. The text data subfolder contains questionnaires and vocabulary tests, both the blank tests, and the tests filled in by participants (Word documents with participant IDs).

Finally, the folder Analysis will contain the Statistical data (see above). This subfolder will also contain a log file (.txt) giving a detailed description of each step in the data cleaning, pre-processing, and analysis procedure, as well as the accompanying R codes and reference to the relevant files and where these are stored.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

To the best of my knowledge, there is currently no metadata standard available for the type of data that will be collected as part of the project.

## 3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on the university's central servers via OneDrive.

How will the data be backed up?

OneDrive has automatic daily back-up procedures.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

NA

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Data will be stored on the university's central servers via OneDrive. OneDrive is a secure university-wide platform, which should prevent access by unauthorized persons.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

Use of the university's central servers is free of cost. No additional data storage will be necessary, since the university's central servers contain enough space for all the data generated in the project.

#### 4. Data preservation after the end of the research project

**Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).**

Personal data will be stored on the university's central servers (with automatic back-up procedures) for 10 years, after which the data will be deleted.

**Where will these data be archived (stored and curated for the long-term)?**

The university's central servers.

**What are the expected costs for data preservation during the expected retention period? How will these costs be covered?**

NA (see above).

#### 5. Data sharing and reuse

**Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.**

- Yes, in an Open Access repository

Materials generated for the project (tests, eye-tracking experiments, etc.) will be made publicly available via the Open Science Framework (OSF). This includes the following data sets:

1. Vocabulary tests
2. Questionnaires
3. Eye-tracking experiments
4. Statistical data (r codes, tables, figures, etc.)
5. Accompanying information

**If access is restricted, please specify who will be able to access the data and under what conditions.**

Results of the study (data generated for individual participants, including test results and eye-tracking results) will only be published in summarized form, full data sets or raw data will not be shared beyond the data sharing contract with Tohoku University. Non-pseudonymized personal data will only be accessible to the researchers (Eva Puimège and Elke Peters, for all data collected in Belgium, or Takumi Uchiyama, for data collected in Japan).

**Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.**

- Yes, Privacy aspects

Personal data (e.g., name, email address, ID, age, gender) will be pseudonymized. Each participant will be assigned a randomized ID code that will be used in all data processing and analysis. Information that can identify a participant (name and email address) will be stored in a single file (.xls) along with the corresponding ID code. After data collection, this file will only be consulted if a participant wants to exercise their right of access to the data. Personal data such as age and gender will only be shared in aggregated form. Personal data that can identify individual participants (e.g., name, email address) or bank details will not be shared.

**Where will the data be made available? If already known, please provide a repository per dataset or data type.**

Open Science Framework (OSF)

**When will the data be made available?**

Data will be made available upon publication of research results.

**Which data usage licenses are you going to provide? If none, please explain why.**

Data will be made available under a creative commons attribution license (CC-BY 4.0), so that users have to give credit to the original data creators.

**Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.**

- Yes

We will add a DOI number to our dataset.

**What are the expected costs for data sharing? How will these costs be covered?**

The OSF is currently free of cost.

## **6. Responsibilities**

**Who will manage data documentation and metadata during the research project?**

Eva Puimège

**Who will manage data storage and backup during the research project?**

Eva Puimège

**Who will manage data preservation and sharing?**

Elke Peters

**Who will update and implement this DMP?**

Eva Puimège