Plan Overview

A Data Management Plan created using DMPonline.be

Title: Extramural English and young learners' development of implicit and explicit second language knowledge

Creator: Joke Van Mol

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Project abstract:

Recently, a growing number of studies have shown the potential of extramural English (= out-of-school engagement with English) for language learning. However, most of these studies have used explicit measures of L2 knowledge, such as multiple choice vocabulary tests and untimed grammaticality judgements. The present proposal aims to advance our understanding of the input-acquisition relationship by investigating the mechanisms underlying explicit and implicit language learning by English-as-Foreign-Language (EFL) learners who are exposed to large amounts of English language input outside of school, but have not started formal English instruction yet. In two correlational studies, we will investigate the relationship between (i) extramural English and (ii) explicit and implicit word and grammar knowledge. Further, in two longitudinal studies, we will study their development of explicit and implicit word and grammar knowledge in English.

ID: 211364

Start date: 01-10-2024

End date: 01-10-2028

Last modified: 17-12-2024

Extramural English and young learners' development of implicit and explicit second language knowledge
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)

This project will generate new data, including personal data:

1. **Questionnaires**: a parental and a child questionnaire made in the Gorilla experiment builder (gorilla.sc), including Likert scales, open questions and multiple choice questions (data type = survey/questionnaire; format = .html)
2. **Questionnaire responses**: spreadsheets containing responses to the questionnaires, generated by Gorilla (data type = observational, format = .csv, .xlsx)
3. **Inhibitory control test**: a Flanker or Stroop task (e.g., Black/White Stroop, Bear/Dragon Stroop), measuring participants' inhibitory control made in the Gorilla experiment builder (data type = online experiment; format = .html)
4. **Inhibitory control test results**: spreadsheets containing results on the Flanker or Stroop task, generated by Gorilla (data type = experimental; format = .csv, .xlsx)
5. **Vocabulary instruments**: a digital version of the Picture Vocabulary Size Test (Anthony & Nation, 2017), a sentence-reading task, a visual-world task, and stock images for the visual-world task (data type = online experiment; format = .html)
6. **Vocabulary test results**: spreadsheets containing results on the PVST, sentence-reading task, and visual-world task generated by Gorilla (data type = experimental; format = .csv, .xlsx)
7. **Grammar instruments**: digital grammar tests which will be designed in the Gorilla experiment builder, including an adaptation of the Grammaticality Judgment Task (Pfenninger, 2011), a word-monitoring task, and a sentence-reading task (data type = online experiment; format = .html)
8. **Grammar test results**: spreadsheets containing results on the GJT and word-monitoring task, generated by Gorilla (data type = experimental; format = .csv, .xlsx)
9. **Eye-tracking experiments**: Experiment Builder files, deployed experiment, interest area files, spreadsheet containing the data source, other small files generated as part of the experiment (data type = software; format = .csn, .xlsx, .txt, .ebd, .exe)
10. **Eye-tracking data**: raw eye-tracking data, data viewer projects, and spreadsheets containing cleaned data (data type = experimental; format = .cvs, .xlsx, .edf, .evs)
11. **Consent form**: parents and children will receive a consent form on paper, which will later be stored in spreadsheets (data type = survey/questionnaire; format = .docx, .csv, .xlsx)
12. **Statistical data**: R codes, JASP files and spreadsheets containing results and aggregated data (data type = compiled/aggregated data; format = .r, .jasp, .csv, .xlsx)

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. Designation of responsible person (If already designated, please fill in his/her name.): Prof. dr. Elke Peters (elke.peters@kuleuven.be)
2. Storage capacity/repository
    - during the research, the collected data will be stored on the university's central servers via OneDrive. This platform is used university-wide to store data, and has automatic daily back-up procedures.
    - after the research, personal data will also be stored on the university's central servers (with automatic back-up procedures) for 10 years. After that, the data will be deleted.

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)

For this project, personal data will be collected. These data include demographic information, such as name, month and year of birth, gender, SES, language context, and engagement with English activities.

In order to protect the participants' privacy, the collected data will be pseudonymized. Each participant will be assigned a unique code. Their name, class group, school, province, and other personal data that can identify a participant, will be stored along with that unique code in a key code. Hence, those data will not be linked to the other data, unless participants wish to exercise their right of access to the data. The results of the questionnaires and tests will be stored in a second document. The PhD student will have access to both the key code and the pseudonymized dataset. The PI and co-supervisor will only have access to the pseudonymized dataset.

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

NA

Extramural English and young learners' development of implicit and explicit second language knowledge
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 digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Questionnaires	A parental and a child questionnaire made in the Gorilla experiment builder (gorilla.sc), including Likert scales, open questions and multiple choice questions	Generate new data	Digital	Other: online survey/questionnaire	.html	<1GB	
Questionnaire responses	Spreadsheets containing responses to the questionnaires, generated by Gorilla	Generate new data	Digital	Observational	.csv, .xlsx	<100GB	

Inhibitory control test	A Flanker or Stroop task (e.g., Black/White Stroop, Bear/Dragon Stroop), measuring participants' inhibitory control made in the Gorilla experiment builder	Generate new data	Digital	Other: online survey/experiment	.html	<1GB	
Inhibitory control test results	Spreadsheets containing results on the Flanker or Stroop task, generated by Gorilla	Generate new data	Digital	Observational/experimental	.csv, .xlsx	<100GB	
Vocabulary instruments	A digital version of the Picture Vocabulary Size Test (Anthony & Nation, 2017), a sentence-reading task, a visual-world task, and stock images for the visual-world task	Generate new data	Digital	Other: online survey/experiment	.html	<1GB	
Vocabulary test results	Spreadsheets containing results on the PVST, sentence-reading task, and visual-world task generated by Gorilla	Generate new data	Digital	Observational/experimental	.csv, .xlsx	<100GB	
Grammar instruments	Digital grammar tests which will be designed in the Gorilla experiment builder, including an adaptation of the Grammaticality Judgment Task (Pfenninger, 2011), a word-monitoring task, and a sentence-reading task	Generate new data	Digital	Other: online survey/experiment	.html	<1GB	

Grammar test results	Spreadsheets containing results on the GJT and word-monitoring task, generated by Gorilla	Generate new data	Digital	Observational/experimental	.csv, .xlsx	<100GB	
Eye-tracking experiments	Experiment Builder files, deployed experiment, interest area files, spreadsheet containing the data source, other small files generated as part of the experiment	Generate new data	Digital	Software	.csv, .xlsx, .txt, .ebd, .exe	<100GB	
Eye-tracking data	Raw eye-tracking data, data viewer projects, and spreadsheets containing cleaned data	Generate new data	Digital	Experimental	.csv, .xlsx, .edf, .evs	<100GB	
Consent form	Parents and children will receive a consent form on paper, which will later be stored in spreadsheets	Generate new data	Digital	Other	.docx, .csv, .xlsx	<1GB	
Statistical data	R codes, JASP files and spreadsheets containing results and aggregated data	Generate new data	Digital	Other	.r, .jasp, .csv, .xlsx	<100GB	

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

The project will soon be submitted to PRET for ethical approval (G-2024-8537).

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

For this project, personal data will be collected, more specifically through the parental and child questionnaires. These data include demographic information, such as name, month and year of birth, gender, SES, language context, and engagement with English activities.
In order to protect the participants' privacy, the collected data will be pseudonymized. Each participant will be assigned a unique code. Their name, class group, school, province, and other personal data that can identify a participant, will be stored along with that unique code in a key code. Hence, those data will not be linked to the other data, unless participants wish to exercise their right of access to the data. The results of the questionnaires and tests will be stored in a second document. The PhD student will have access to both the key code and the pseudonymized dataset. The PI and co-supervisor will only have access to the pseudonymized dataset.

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.

- No

NA

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 collected data will be stored in a project folder, of which the name will contain the project ID. Subfolders will be created for the different studies and, within those folders, separate folders will be created for:

1. Information letter (.docx), information presentation (.pptx), and informed consent forms for the participants and for their parents (.docx)
2. Questionnaires, vocabulary instruments, grammar instruments
3. Eye-tracking data: containing a .txt file (describing what the data represent and how they were generated) and the lab book (describing lab setup, eye tracker settings, host and display settings, as well as the data collection procedure)
4. Test results
5. Statistical data (containing a log file (.txt) with a detailed description of the data cleaning, pre-processing, and analysis procedures

The project folder will also contain a codebook, with all the information necessary to reproduce and use the data (e.g., study design, methodology, data coding and analysis).

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 our 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?

We will use OneDrive, since it 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

KU Leuven staff members have 250GB of storage space on OneDrive.

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?

Since we will be using OneDrive for data storage (which is free of cost), there will likely be no additional costs.

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 through OneDrive 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

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 that will be used for this project (questionnaires, vocabulary and grammar instruments, eye-tracking experiments, statistical data and accompanying information) will be made publicly available in an Open Access repository.

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

Results of the study that will be made available in Open Access, will be published in summarized form. Full data sets will only be accessible to the PI, co-supervisor and PhD student. Only the PhD student has access to the non-pseudonymized data.

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

Participants' personal data (e.g., name, gender, age) will be pseudonymized. Each participant will be assigned an ID code, which will be stored separately in a key code, along with other information that can identify participants. This file will only be consulted when participants want to exercise their right of access to the data.

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

Data will be made available in the Open Science Framework (OSF, https://osf.io/), and the IRIS database (https://www.iris-database.org/). More specifically, the OSF will be used to publish the data in summarized form and R codes, whereas the IRIS database will be used to publish the test materials.

When will the data be made available?

Data will be made available upon publication of the 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?

Both OSF and IRIS are currently free of cost.

6. Responsibilities

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

Joke Van Mol

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

Joke Van Mol

Who will manage data preservation and sharing?

Elke Peters

Who will update and implement this DMP?

Joke Van Mol