To logic or not to logic in mathematics education? A mixed-methods investigation towards evidence-based instruction of mathematical proof. Application DMP

## Questionnaire

The questions in this section should only be answered if you are currently applying for FWO funding. Are you preparing an application for funding?

No

To logic or not to logic in mathematics education? A mixed-methods investigation towards evidence-based instruction of mathematical proof.	
DPIA	

DPIA

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

Question not answered.

GDPR	gic of not to logic in mathematics education: A mixed-methods investigation towards evidence-based instruction of mathematical proof.
GDPR	
Have y	ou registered personal data processing activities for this project?

Question not answered.

To logic or not to logic in mathematics education? A mixed-methods investigation towards evidence-based instruction of mathematical proof. 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.

Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume Explanation (MB/GB/TB)	Digital Data Volume Estimate (in GB)	Physical volume (in pages)
01pilots	O1: qualitative analysis of pupil answers on math test; first piloting test with some pupils	Newly generated	Physical: paper-notes	/				100
O1interviews	O1: qualitative analysis of pupil answers on math test; in next piloting round also including interviews about the test	Newly generated	Digital: video recording	Observational	.mp4	<100 GB (30 interviews of each hour at 720p)	100	
O1largecollect	O1: qualitative analysis of pupil answers on math test; the final test is distributed to large sample (pupils fill in test on paper, a scan automatically converts this in KU Leuven question platform)	Newly	Physical: paper-notes, then Digital in used online question platform	Observational		<100 GB (no restriction from ANS, essentially 300*20*2MB per image)		1800
O1encoding	01: qualitative analysis of pupil answers on math test; coding of pupil answers	Newly generated	Digital: text- labeling of pupil answers	Own Creation	nVivo - object	<100 MB	0.1	
02interviews	02: quantitative study on correlation of several measures (math test, logical reasoning,); open-ended math test is converted to closed-endet test using pupil interviews	Newly generated	Digital: video recording	Observational	.mp4	<100 GB (30 interviews of each hour at 720p)	100	
O2materials	02: quantitative study on correlation of several measures (math test, logical reasoning,); all the final testing instruments together can also be considered as research data		Digital: code and possibly images/videos	Own Creation	tex/.pdf for LaTeX; .png for contained images; .html/.css/.js for web-version; Moodle object for alternative web-version (testing instruments will be implemented in 3 versions: LaTeX for printing on paper, Moodle for integration with learning interventions and Javascript for standalone in other research projects)	2100 MR	0.1	

O2largecollect	02: quantitative study on correlation of several measures (math test, logical reasoning,); collection of pupil answers on large scale	Newly	Digital: database of multiple choice answer, reaction time, variables	Observational	.csv	<100MB	0.1	
03materials	O3: Design digital learning intervention improving math score; the final learning materials are also research data	Newly generated	Digital: code and possibly images	Own Creation	.tex/.pdf for LaTeX; .ggb for GeoGebra files (standard program for geometry exercises); .mp4 for videos; .png for images		1	
03moduleinteract	O3: Design digital learning intervention improving math score; Pupils will interact with both an online learning environment, generating a data set (scores on exercises, duration spent on exercises,) and will complete some on-paper exercises.	Newly generated	Digital: database in online learning platform + Physical: paper-notes of participating pupils	Observational	.csv (after export from Moodle)	<100 MB	0.1	
03fields	O3: Design digital learning intervention improving math score; Own field notes while observing the classroom during intervention	Newly generated	Physical: paper notes	Own Creation				100
03classrecording	O3: Design digital learning intervention improving math score; Classroom recording after the intervention to verify ideas for improvement O4: Randomized-	generated	Digital: video recording	Observational	.mp4	<100 GB (each hour at 720p for max 10 hours)	30	
O4moduleinteract	controlled trial for comparing effectiveness of several different interventions; pupils interactions in the online learning environment generates a data set, same as O3moduleinteract	Newly generated	Digital: database in online learning platform	Observational	.csv	<100 MB	100	
O4testings	04: Randomized- controlled trial for comparing effectiveness of several different interventions; large scale collection before and after the interventions, using testing instruments from 02 (so same as 02collect)	Newly	Digital: database of multiple- choice answer, reaction time, variables	Experimental ("controlled observation of humans")	.csv	<100MB	100	
overall-statsanalysis	All code for quantitative statistical analyses (02, 03, 04)	Newly generated	Digital: code	Own Creation	.R	<100 MB	10	

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:

Created using DMPonline.be. Last modified 27 March 2025

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

This project already received ethics approval from the SMEC committee with application number G-2024-8859.

This project is in the area of mathematics educational research, and thus needs to collect educationally relevant data from pupils in order to investigate teaching effectiveness. Participants are informed in full in advance, their experience during participation is almost indistinguishable from ordinary mathematics lessons (filling in a test or completing an online learning module) + parental consent is required in advance.

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

This project already received personal data processing approval from the GDPR-unit in KU Leuven with application number G-2024-8859.

In short, we only collect essential personal data of the pupil (name, gender, age), of the class (study program, study year) and of the school (which school group), together with student reponses to mathematical testing instruments and control instruments (for example socio-economic status). Before participation, parents have the time to go through all information in an informed consent and pupils themselves also have the option to opt out at any stage in the research process. All personal data is pseudonymized immediately after collection, and the pseudonymization keys are stored separately and only kept for the minimally needed time.

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.

Yes

As "open-source" is one of this project's spearheads, the created testing instruments (O2materials) and the created educational intervention online modules (O3materials) will be released publicly and easily copy-able. This means anyone in the world (for example researchers wanting to test the same skills or textbook authors in proving) are allowed to build on the produced materials, including commercialisation (for example selling derived textbooks to students). We are purposely allowing commercial derived works, to increase the impact of our research project on everyday practice (which is mainly guided by textbooks from commercial publishers).

Commercial valorization of the datasets containing student answers is neither appropriate nor wanted

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

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.

• Yes

Upon verifying with KU Leuven research support (LRD), copyright for the created testing instruments (02materials) and the created educational intervention online module (03materials) will fall automatically on the creator (ie: the PhD-student of this project). More details later in the question on which license is going to be used for publicly sharing these materials for teachers and other researchers.

Other listed datasets do not pose specific legal issues.

### 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).

I reused the table of before as there are many different types of data (since this is a mixed-methods project). In future questions I will use a spreadsheet again if the answer is different for each, but will stick to paragraphs for brevity if the same applies to many datasets.

Dataset Name	Description	Documentation	Metadata
O1pilots	O1: qualitative analysis of pupil answers on math test; first piloting test with some pupils	Documentation is an internal .md file explaining the design decisions, choice of questions, instructions on how it was used,	/
01interviews		Documentation is an internal .md file containing the composed interview questions + (if further analysed) generated documentation from coding in NVivo	File-properties within .mp4-file + (if further analysed) generated metada from NVivo
O1largecollect	test; the final test is distributed to large sample (pupils fill in test on paper, a scan	Finetuned internal .md file explaining design choices etc as before (later turned into accompanying paper) Digitized on-paper exams will be stored on pre-existing question platform ANS.app which has their own documentation on how to use it (https://support.ans.app/hc/en-us/sections/16771659641233-Print-scan). Data will not be shared with external parties on ANS (ie, internal .md file is enough)	Metadata standard from ANS.app is used (see https://support.ans.app/hc/en-us/articles/7652874788369-Open-question-for-written-assignments) Export of questions possible in QTI-format (industry standard) (https://support.ans.app/hc/en-us/articles/360020000158-Export-options-for-exercises-questions-and-insights), the pupil answers are scanned into pdf's when digitizing

O1encoding	analysis of pupil answers on math test; coding of	The used coding scheme will be implemented on ANS.app (as a custom "grading scheme", see ANS own documentation (https://support.ans.app/hc/en-us/articles/4406832054929-Add-a-grading-scheme-to-your-questions)) with a manually updated .csv file as copy of the coding scheme with its explanation	Metadata standard from ANS.app is used (see https://support.ans.app/hc/en-us/articles/7652874788369- Open-question-for-written-assignments)
O2interviews	reasoning,);	Documentation is an internal .md file containing the composed interview questions + (if further analysed) generated documentation from coding in NVivo	File-properties within .mp4-file + (if further analysed) generated metada from NVivo
02materials	O2: quantitative study on correlation of several measures (math test, logical reasoning,); all the final testing instruments together can also be considered as research data	The accompanying paper (building on previous .md	No typical meta-data standard exists for jsPsych testing instruments up to my knowledge. We will adapt the one from Pavlovia (the hosting website for standalone experiments) https://pavlovia.org/docs/experiments/create-jsPsych
O2largecollect	correlation of several measures (math test, logical reasoning,); collection of pupil	We will include a README.txt file right next to the saved dataset (already locally, but also later on osf.io) with context explanation of this dataset + link to the shared statistical analysis code on osf.io The accompanying paper (building on previous.md files) will also explain its philosophy (and the README file will contain researcher instructions)	No metadata standards for these variables exist. We will define our own in the associated statistical analysis code (in R), using this package (https://dataset.dataobservatory.eu/articles/metadata.html) This metadata will be manually copied to the dataset on osf.io Dataset and Analysis code will always link to eachother.
03materials	intervention improving math score; the final learning materials	Materials are mostly created in LaTeX, alongside a teacher guide in LaTeX. The associated Moodle course explicitly links to the teacher guide.  An accompanying paper will also include how these materials were constructed for researchers.	We will use standards from oercommons.org and Klascement to make these educational resources easily findable.
03moduleinteract	generating a data set (scores on exercises, duration spent on	This dataset contains both physical and digital data. The digital data is saved in the Moodle course, next to the teacher guide and a link to the accompanying paper.  Apart from that, standard Moodle plugins will be used that already have their own documentation (such as https://docs.moodle.org/405/en/Assignment_activity) The on-paper exercises will be saved together with a printed version of the .md files describing the study.	No metadata standards for these variables exist. We will define our own in the associated statistical analysis code (in R), using this package (https://dataset.dataobservatory.eu/articles/metadata.html) after export from Moodle to a local .csv file
03fields	O3: Design digital learning intervention improving math score; Own field notes while observing the classroom during intervention	/	/

O3classrecording	O3: Design digital learning intervention improving math score; Classroom recording after the intervention to verify ideas for improvement		Minimal, only file-properties within .mp4-file
O4moduleinteract	O4: Randomized- controlled trial for comparing effectiveness of several different interventions; pupils interactions in the online learning environment generates a data set, same as O3moduleinteract	Same as O3moduleinteract without paper exercises	No metadata standards for these variables exist. We will define our own in the associated statistical analysis code (in R), using this package (https://dataset.dataobservatory.eu/articles/metadata.html) We will manually copy the metadata to osf.io.
O4testings	O4: Randomized- controlled trial for comparing effectiveness of several different interventions; large scale collection before and after the interventions, using testing instruments from O2 (so same as O2collect)	Same as O2largecollect	Same as O2largecollect
overall- statsanalysis		During active development, a README file + comments in code will document the code in same folder (on Gitlab KU Leuven). The documentation will also be included when sharing the statistical analysis files on osf.io	As described before, the .R code itself will not be tagged with metadata, but it will show what metadata was added to the dataset and link to the dataset on osf.io

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.

• Yes

See the previous table.

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

## Where will the data be stored?

Dataset Name	Description	Where data stored?	Backed up?	Securely stored and not accessed or modified by unauthorized persons?	Expected costs for storage/backup? Who covers it? (Shai drive: https://icts.kuleuven.be/sc/english/storage/desktc file-storage#hoeveel)
01pilots	O1: qualitative analysis of pupil answers on math test; first piloting test with some pupils	Closet in the researcher's private office	Digital scan on KU Leuven network drive	Both closet and office are locked + office inconspicuously looks like broom closet	None, closets present
O1interviews	niloting round	Shared KU Leuven drive (J: drive) (Not on personal OneDrive as this cannot be pseudonimyzed)	Automatically daily		around 40 euro's total (as we will only store this for one year), covered by existing group budget for network share or FWO credit

O1largecollect	answers on math test; the final test is distributed to large sample (pupils fill in test on paper, a scan automatically	Informed consents on paper in closet (as before). Physical mathematics tests in closet as well. Scanned pdf's with pseudo-id on OneDrive (link between id's and pseudo-id's is stored by data manager of research group CIP&T (i.e., Marc Vlecker, ithrsy://ppw.klueven.be/ppw-dict/dictservicedesk/where-are-my-network-drives)). A scanned version is stored on the Dutch company ANS.app which has already an agreement with KU Leuven to be used as question platform for its students.	Monthly backup to personal hard drive of data on OneDrive; ANS handles own internal backups	Closet (see before). OneDrive and ANS are both authenticated through KU Leuven accounts	There are no extra costs: OneDrive is already paid for PhE researcher regardless of usage. ANS has existing agreem with KU Leuven to host student data (after communicatio with ICTS our expected 300 students are allowed within tiagreement)
O1encoding		the Dutch company ANS.app which has already an agreement with KU Leuven to be used as question platform for its students.	ANS handles own internal backups + manually updated .csv file	Authentication through KUL, limited to researchers collaborating on the encoding	None extra, see above
02interviews	O2: quantitative study on correlation of several measures (math test, logical reasoning,); open-ended math test is converted to closed-ended test using pupil interviews		Automatically daily by ICT KU Leuven	Authentication by KUL-account, shared with exclusively promotors for collaborative analysis	around 40 euro's total (as it is only stored for one year), covered by existing group budget for network shares or F credit
02materials	O2: quantitative study on correlation of several measures (math test, logical reasoning,); all the final testing instruments together can also be considered as research data		Version control Git is used	Authentication happens through KUL accounts; project settings on Gitlab allow fine-grained control of what is visible	None, creating project is free (size is minimal: only code I documents)
O2largecollect	O2: quantitative study on correlation of several measures (math test, logical reasoning,); collection of pupil answers on large scale	Informed consents on paper in closet (as before). Informed assent by pupils stored within collected digital dataset (and undergoes same pseudonimyzation) Link between id's and pseudo-id's is stored by data manager of research group CIP&T (i.e., Marc Vlecker; https://ppw.kuleuven.be/ppw-dict/dictservicedesk/where-are-my-network-drives). Experiment is hosted on Pavlovia.org (all servers located within EU and GDPR-compliant, information on https://pavlovia.org/docs/home/ethics; used several times before at Psychology faculty KU Leuven). After experiment completion, pseudonimyzed dataset is exported to own OneDrive (and Pavlovia project with data deleted: there are better places for sharing used datasets)	Monthly backup to personal hard drive of data on OneDrive	Same for closet as before. Pavlovia only accessible by password- protected account	None, KU Leuven includes Pavlovia license for researcher
O3materials	score; the final	LaTeX-files are maintained on Github repository. Own Moodle server will be set up on KU Leuven hardware for hosting learning environment as a whole (including videos for example)	Manual monthly backup (https://docs.moodle.org/405/en/Site_backup) to OneDrive by researcher	Authentication through KUL, limited to researchers collaborating	100 euro per year (every year) for Moodle server, covered FWO credit (https://icts.kuleuven.be/sc/publicatie/webhosting#hoev Github free.

O3moduleinteract	O3: Design digital learning intervention improving math score; Pupils will interact with both an online learning environment, generating a data set (scores on exercises, duration spent on exercises,) and will complete some on-paper exercises.	Digital interactions on own Moodle Server on KU Leuven hardware. Physical notes in closet. (Pseudonimyzed grades then exported to OneDrive after completion of collection (https://docs.moodle.org/405/en/Grade_export))	Manual monthly backup (https://docs.moodle.org/405/en/Site_backup) to OneDrive by researcher	Authentication through KUL, limited to researchers collaborating. Students will receive random id + password to participate on the Moodle (and the link between random id and actual name is stored with data manager of CIPT, see before)	Max 10 euro per year (see link above)
O3fields	notes while observing the classroom during intervention	Closet in the researcher's private office	/	Both closet and office are locked + office inconspicuously looks like broom closet	None, closets present
03classrecording	score; Classroom recording after the intervention to verify ideas for improvement	pseudonimyzed)	Automatically daily by ICT KU Leuven	Authentication by KUL-account, shared with exclusively promotors for collaborative analysis	around 15 euro's total, covered by existing group budget t network shares or FWO credit
	O4: Randomized- controlled trial for comparing effectiveness of several different interventions; pupils interactions in the online learning environment generates a data set, same as O3moduleinteract	Same as O3moduleinteract without paper exercises	Same as 03moduleinteract without paper exercises	Same as O3moduleinteract without paper exercises	Same as O3moduleinteract without paper exercises
04testings	O4: Randomized- controlled trial for comparing effectiveness of several different interventions; large scale collection before and after the interventions, using testing instruments from O2 (so same as O2collect)		Same as O2largecollect	Same as O2largecollect	Same as O2largecollect
overall-	All code for quantitative statistical analyses (O2, O3, O4)	Personal laptop + Gitlab KU Leuven for version control	Version control Git is used	Authentication happens through KUL accounts; project settings on Gitlab allow fine-grained control of what is visible	None, creating project is free (size is minimal: only code l documents)

### How will the data be backed up?

See overview before.

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

 $\label{thm:condition} \textbf{Total physical data is less than 3000 pages, which can fit in the mentioned closet.}$ 

Digital data in total (across all platforms) is limited to 250 GB, most of which comes from video materials which will be stored on the Shared KU Leuven drive of the research group. Other online platforms (ANS, Pavlovia, Gitlab, ...) either do not have exact data limits (as typical usage is less than 10 Mb's) or the expected data size is covered by an existing KU Leuven agreement.

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

To summarize the table before: physical data is stored in a locked closet within the researcher's personal, locked office.

Digital data is stored on KU Leuven server's or on reliable platforms the university has existing agreements with. All authentication happens through KU Leuven's central login.

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

See the overview before.

Expected costs are well below the available FWO credit for yearly research expenses (on the scale of 150 euro per year).

### 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...).

Wherever possible, data will be retained for at least five years. Exceptions are the created videos for the interviews, as these are hard to pseudonimyze and will be converted into written transcriptions (which will be kept). Also the field notes provide little relevant information for later re-use and will thus not be stored long-term. And lastly, the dataset of interactions in the online learning platform (during development, so 03 is of little use to other researchers)

The binding key (connecting pseudo-ids and identifying information) will be stored only for limited time and then securely destroyed.

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

Physical data will be archived in the research group archives for at least 5 years.

Digital data will be archived (and shared, see later) on the Zenodo platform. Source code will be made public on the Gitlab from KU Leuven. The designed educational resources will be deposited in oercommons.org and (flemish database of such resource) "Klascement".

For integration with the FWO's automatic checks on publications, we will also register a record in Lirias (without the data itself) to afford integration with the FRIS research portal.

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

None

#### 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
   Yes, in a restricted access repository (after approval, institutional access only, ...)

All data with re-use value will be shared (see the exceptions in the data retention section). The testing materials used will be shared in a restricted access repository, to prevent students from preparing their assessment in advance

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

Access is only restricted to the used testing instruments. Access is granted by contacting the researcher or their promotor (after enddate project) by email.

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

As mentioned before in the data preservation sections, video material in O1, O2 and O3 is hard to pseudonimyze and provides little re-use value.

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

See the question on where data will be archived.

## When will the data be made available?

Data of the corresponding study will be made available after completion of that study within the project.

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

After consulting the Research support desk from KU Leuven (LRD) about copyright and licensing, we have made a license selection for all shared data-types.

Dataset Name	Which data usage licenses are you going to provide? If none, please explain why.
O1pilots	/
O1interviews	/
O1largecollect	CC BY- NC-SA
01encoding	CC BY- NC-SA
O2interviews	/
02materials	GNU GPLv3
O2largecollect	CC BY- NC-SA
03materials	CC BY- SA
O3moduleinteract	
O3fields	/
03classrecording	/
04moduleinteract	CC BY- NC-SA
O4testings	CC BY- NC-SA
overall-statsanalysis	GNU GPLv3

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

The used repositories automatically assign a DOI (except for oercommons for the educational materials of O3, this is not a problem as the designed materials are made to be repeatedly adjusted)

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

None, Zenodo has a free limit of up to 50 GB which is well over our expected data size.

### 6. Responsibilities

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

Alexander Holvoet (PhD-student of this project)

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

Alexander Holvoet (PhD-student of this project); DICT (IT-department psychology)

Who will manage data preservation and sharing?

Alexander Holvoet (PhD-student of this project); Wim Van Dooren (promotor, after project ends)

Who will update and implement this DMP?

Alexander Holvoet (PhD-student of this project)