Plan Overview

A Data Management Plan created using DMPonline

Title: Developing a ready-made pedagogical framework to support the long-term academic writing skills of university students in Estonia and the US

Creator: Roger Yallop

Principal Investigator: Roger Yallop

Project Administrator: Djuddah Leijen

Affiliation: University of Tartu

Funder: European Research Council (ERC)

Template: ERC DMP

ORCID iD: 0000-0003-2240-894X

Project abstract:

This project furthers the research described in my PhD thesis where I designed a pedagogical framework to support the long-term writing skills of L2 English PhD students in an Estonian context. The framework originates from socio-cultural theory and describes how students can more effectively give and receive written feedback in their discipline-specific writing groups as they develop a better sense of community. I will conduct large-scale replication studies in a mixed-method research design to adapt my framework to other teaching and socio-cultural contexts. Data are obtained from a wide range of writing groups (e.g., different study levels and disciplines) in naturalistic teaching settings from US and Estonian universities, and analysed using machine learning algorithms derived from grounded theory. The findings will benefit writing pedagogy by providing a refined framework that can be tweaked to improve students' long-term writing skills in both face-to-face and online mediums.

In total there are seven planned studies as follows:

- ***Study 1.*** The influence of '*synchronous writing group meetings*' on the peer feedback process.
- ***Study 2.*** The influence of *direct teaching instruction*' on the peer feedback process
- ***Study 3.*** The influence of '*individual writing styles and strategies*' on the peer feedback process
- ***Study 4.*** The effectiveness of written peer feedback comments within US academic writing courses; a large-scale replication study of Yallop & Leijen (2018)
- ***Study 5.*** The effectiveness of written cover letter comments as compared to writing assessment rubrics within US academic writing courses; a large-scale replication study of Yallop & Leijen (2020).
- ***Study 6***. The affect and effect of asynchronous written feedback comments on the peer feedback process: An ethnographic case-study approach within one L1 American doctorate writing group; a replication study of Yallop & Leijen (2020).
- ***Study 7*****.** Developing a flexible writing pedagogical framework to promote the peer feedback process in writing groups, a validation study of Yallop (2021: 105-110).

Studies 1-3 mainly collect data from participants at Tartu University (TU). Studies 4-6 mainly collect data from participants at the Ohio State University (OSU). Study 7 mainly conducts secondary analysis on data collected from the participants who participated in Studies 1-6. Thus, Study 7 contains datasets collected from both TU and OSU.

As such, this Data Management Plan (DMP) addresses how I will manage data obtained from two different countries (Estonia and the USA), and how the ethical and legal issues will be addressed.

ID: 100796

Start date: 01-07-2022

End date: 30-06-2024

Last modified: 14-12-2022

Grant number / URL: PUTJD1121

Copyright information:

Developing a ready-made pedagogical framework to support the long-term academic writing skills of university students in Estonia and the US

Summary

Project Acronym

DWPF

Project Number

PUTJD1121

Provide a dataset summary

1. Data Collection
What data will you collect or create?
Background
The project will mainly collect primary data from human adult participants. The participants will be predominantly students participating on academic writing courses at universities in both Estonia (mainly the University of Tartu) and the USA (mainly the Ohio State University). These students will produce written artefacts (e.g., draft texts, feedback cover letters, and feedback comments) as a natural consequence of their participation on the course. This study aims to collect such student-produced artefacts as well as their perceptions on why they produced such artefacts through qualitative research methods (e.g., through surveys, one-to-one interviews, and focus group meetings). In addition, there will be one dataset collected from writing instructors based in Europe and the USA eliciting their perceptions about the quality of such student-produced artefacts through an online survey. As human participants are involved in this study, there will be additional data sets created in obtaining their intention to participate through an online survey (University of Tartu), and then obtaining their signed informed consent and in the subsequent pseudonymising of their data (University of Tartu and the Ohio State University). Datasets collected from students at the University of Tartu are subject to the conditions of the ethical approval granted to this research project as stipulated by the Research Ethics Committee of the University of Tartu. Similarly, datasets collected from students at the Ohio State University are subject to the conditions of the ethical approval granted to this research project as stipulated by the Internal Revenue Board of the Ohio State University (OSU).

There are differences between the two ethical committees on how informed consent can be obtained from the participants, and this results in different datasets. As stipulated by the Research Ethics Committee of the University of Tartu, potential participants cannot be approached directly by their responsible lecturers and/or the principal investigators and project collaborators as this may be considered undue force to participate in this 'clinical trial' as stated on the ethical approval form. As such, potential participants are invited to complete an online survey to signal their intention to participate. If the potential participant signals their intention to participate in this research project, then they will be invited to sign (either by hand or digitally) an informed consent form to participate in this research. Thus, this results in three datasets: (i) the potential participants' intention to participate through an online survey, (ii) hand-written informed consent forms, (iii) digitally signed informed consent forms and scanned hand-signed informed consent forms. Conversely, and under the rules of the US Internal Review Board, it is normally allowed to approach the students directly to participate in this research project as it is for the common good. Furthermore, this research project is not likely to be classified as a 'clinical trial', and as such the ethical rules are much more relaxed. As such, the only datasets obtained from the participants are hand-written and/or digitally signed ethical consent forms.

In addition, any project collaborating researcher outside the University of Tartu ('External Researcher') is required to sign a 'Researcher Authorisation to Use Data form' in order to analyse pseudonymised data obtained from the participants from the University of Tartu. It is not envisaged that these same requirements are demanded from the Internal Revenue Board of OSU. All participant original personal data is pseudonymised. This pseudonymisation process results in further datasets that need to be managed. Finally, there are datasets that will be obtained directly from the participants (e.g., student-produced written artefacts) that will constitute the instruments for analysis for this research project. There may also be some analytical datasets produced from the participants from the University of Tartu, but not from participants from OSU (and vice-versa), but for simplicity it is assumed that these datasets are produced by participants from both universities.

Data Generation
All e-datasets generated in Estonia will be stored on UT's server (for electronic documents) and all hardcopies will be stored in the office shared by the principal investigator (Roger Yallop) and the main collaborator (Djuddah Leijen) in a locked drawer (Lossi 3- 306, Tartu, Estonia). Similarly, all e-datasets generated in Estonia will be stored on OSU's server (for electronic documents) and all

hardcopies will be stored in the office shared by the principal investigator (Roger Yallop) and the main external collaborator and Roger's postdoctoral supervisor (Susan Lang) in a locked drawer (4132 C Smith Lab, 174 W. 18th St., Columbus, Ohio, USA). The exact nomenclature for the folders and files will be established on an ad hoc basis.

Thus, the following data will be generated during the project:

1. *Participant Intention to Participate Surveys.* Potential participants (the University of Tartu only) will signal their intention to participate through an online survey (LimeSurvey). The resulting data will be available on the survey's database. As such, no identifying features are assigned to this dataset.

1. *Participant Informed Consent Forms and Researcher Authorisation to use Data Forms.* These forms are either digitally or hand signed by the Participant or External Collaborator and the Principal Investigator (Roger Yallop), the main Collaborating Researcher (Djuddah Leijen), or the main external collaborator (Susan Lang). All hand-signed informed consent forms are scanned as PDF documents. All digitally signed informed consent forms are created as PDFs and digitally signed using appropriate software as according to the requirements of UT or OSU. The resulting digitally signed e-documents are password-protected and stored on UT or OSU's server, and the resulting hand-signed paper PDF documents are secured in a locked drawer in the Principal Investigator (PI) and Main Internal or External Collaborator's office at UT or OSU. All informed consent forms as e-documents are stored under the Master Folder entitled 'ICForms_Est' for all data collected in Estonia and stored on UT's server, and similarly the master folder is entitled 'ICForms_USA' for all data collected in the USA and stored on OSU's server. All subfolders will contain an abbreviation of the course and date, and Study number (e.g., S1, S2, etc.) from which the data was obtained (e.g., 'ComSci2022'). Each participant informed consent form will then be stored in each subfolder using a string of numbers and/or letters (see below identification code lists) to protect the participants' anonymity. All electronic Researcher Authorisation to use Data Forms are stored in one folder entitled 'RADForms' on UT's server only, and each electronic Researcher Authorisation to use Data Form will contain a string of numbers and/or letters (see below identification code lists).

1. *Identification Code Lists.* In the pseudonymisation of data, all interim drafts (mainly as MS Word and MS Excel documents) will be password-protected and stored on UT's server or OSU's server depending on whether the dataset was obtained in Estonia or the USA. So that the data can be connected to the same participant's artefacts, and to the applicable course, each participant's data will be given an identification code as a string of letters and numbers. The identification code to all the Participants' written and aural artefacts will be compiled electronically as an MS Word and/or MS Excel document, and stored on the UT or OSU's server, and a back-up copy on a hard disk will be stored in a locked drawer in the Researcher(s) office (Lossi 3-306) at UT for Estonian data, or in the Researcher's other office at OSU for US data. The main folder will be entitled 'ICL_Est' or 'ICL_USA' for Estonian and USA data respectively (and also stored on UT or OSU's server). All subfolders will contain an abbreviation of the course and date from which the data was obtained (e.g., 'ComSci2022'). Each subfolder will then contain, as a Word document, the participant's name and corresponding pseudonymisation code as a string of letters and numbers that will connect the participant to all their generated artefacts (e.g., draft texts) as well as to their course and other personal data revealed (name, age range, gender, academic writing course, participant discipline, writing language) (see subsection below).

1. *Participant Background Detail Forms.* After giving their informed consent, the Participants are asked to provide a few background details about themselves (e.g., name, gender, age-range, discipline, and course) through a University approved survey environment (e.g., LimeSurvey or Qualtrics). The participants can also decide not to disclose such personal data. This personal data is needed as such variables (e.g., age and gender) can influence the study findings and, thus, these variables will need to be reported statistically (e.g., the group consisted of 2 females and 3 males ...) in publications to ensure the study's credibility. This data is used to link the Participants' artefacts in the pseudonymisation of data of which only the Researchers have access. A back-up copy on a hard disk will be stored in a locked drawer in the Researcher(s) office at UT or at OSU. The Participants' personal data will be destroyed no later than five years after the start of this research project (i.e., by 1st October 2027 at the latest). This personal data (if supplied) will be added to the pseudonymised identification code lists as describe in subsection 4 above. The following nomenclature will be used: gender (G_) (m=male, f=female, u = undisclosed), age (A_) (18-21, 22-25, 26-29, 30-34, 35-39, 40+, u=undisclosed), L1 (L1_) (Est=Estonian, Finn=Finnish, Lat=Latvian, Sw=Swedish, Rus=Russian, Eng=English, Oth = Other) (in Estonia) or (Eng=English, Sp=Spanish, Oth = Other) (in USA): Discipline (D_) (SS=Social Sciences, FS=Formal Sciences, Hum=Humanities, STEM= Technology and Engineering)

1. *Student-produced written artefacts* (mainly participant draft texts, feedback comment letters, feedback cover letters, and revision plans). Depending on the course design and student preferences, these documents will mainly be produced as MS Word files or PDFs. The written artefacts produced by the participants as a natural consequence of their participation on the course will be uploaded to UT's server (e.g., 'Nextcloud' or 'Moodle') or OSU's server (e.g., 'OneDrive' or 'Canvas'). For the analysis, these written artefacts will be pseudonymised and converted into suitable file forms for analysis (e.g., as MS Excel files, Word files, or Text Files as appropriate). The Master Folder is entitled 'StArt_Est' for all data collected in Estonia and stored on UT's server, and similarly the master folder is entitled 'StArt_USA' for all data collected in the USA and stored on OSU's server. The subfolders are entitled with the study number 'StArt_CourseYear_Country, the subsubfolders are entitled by each applicable writing group within the course, and then the student artefacts are stored within these subsubfolders entitled by feedback round according to type of artefact (submitted draft text, feedback cover letter, reviewer cover letter, revision plan, and any other artefact produced as to the course design). The files containing these artefacts will be named using the pseudonymised naming code using an applicable nomenclature to be developed.

1. *Online surveys.* Electronic questionnaires are created in a UT (e.g., LimeSurvey) or OSU (e.g., Qualtrics) approved survey environment that will collect written, aural, and/or visual data as recorded screencast (e.g., a screencast from the participant of their text draft together with their corresponding peer feedback comments, but not visual data of the Participant themselves) from the Participants. Participants will also include writing instructors who give their perceptions on the quality of feedback comments for Study 4. The surveys will be configured so that no email or IP addresses are collected from respondents. For the analysis, the Researcher(s) will only use visual data collected as a screencast to connect the Participant's responses to their aural data and written data. Aural data will be transcribed into written documents (mainly as MS Word and MS Excel documents). During this process, the researcher will use the Participants' visual data as a screencast and aural data directly from the online survey environment (i.e., the Researcher will not download this data to another electronic source (e.g., their own laptop). The resulting written transcripts obtained from any aural and/or visual data supplied as a screencast, and all other written data collected from the survey environment will be converted into suitable file forms for analysis (e.g., as MS Excel files). The student-produced data will be stored on the applicable survey server (either at UT or OSU as applicable). Data will be exported into an appropriate file form (mainly as .csv files) and saved as either excel or word files (as appropriate). A suitable nomenclature for the master folders and subsequent subfolders, and for the files stored in their respective folder that is based on the logic described above will be devised.

1. *Participant interviews, focus group meetings, and writing group meetings.* Interviews, focus group meetings and writing group meetings are conducted face-to-face and/or online, and captured as audio-recordings using a suitable recording device (e.g., QuickTime Player) and capturing device (e.g., a laptop or mobile phone), and uploaded to UT's (e.g., Nextcloud) or OSU's server. Once uploaded, all copies of the earlier audio-recordings are deleted from whatever source they were recorded from. This aural data will be transcribed into written documents (mainly as MS Word and MS Excel documents). The resulting aural transcripts will be converted into suitable file forms for analysis (e.g., as MS Excel files). A suitable nomenclature for the master folders and subsequent subfolders, and for the files stored in their respective folder that is based on the logic described above will be devised.

Size of datasets

Based on current knowledge, the minimum and maximum number of participants for Studies 1-6 is estimated to be between about 200-400 participants of which it is estimated that between 50-100 of these same participants will complete about two online surveys, participate in individual interviews or focus group meetings (one interview or meeting per participant). Also assuming that data is collected over one to five feedback rounds per course where students will produce on average about three written artefacts per feedback round. In addition, Study 1 also collects audio-recordings of writing group meetings (assume 5-10 recordings) and Study 4 collects writing instructors' perceptions about the quality of feedback through surveys (assume between 10-20 participants). As such, I estimate the total amount of data per dataset to be managed as follows:

- **200 - 400** Participant Informed Consent Forms
- **1800 - 3600** Student written artefacts (draft texts, feedback cover letters, feedback review letters)
- **100 - 200** Online surveys (eliciting participants' perceptions about the peer feedback process)
- **50 - 100** Written transcriptions of individual interviews and focus group meetings
- **5-10** Transcriptions of writing group meetings
- **10-20** Online Surveys (eliciting Writing Instructors' perceptions about the quality of feedback)

(ii) How will the data be collected or created?

The primary datasets will mainly be created by students and possibly faculty members on academic writing courses and/or participating in writing groups on past, present, and future courses at both UT and OSU. In addition, and only for study 4, a primary dataset will be obtained from writing instructors located at both European and US universities through a survey eliciting their perceptions about the quality of written feedback comments. More specifically, the seven datasets that will be collected throughout the study can be further categorised into two main types of datasets as follows:

Datasets 1-3. These datasets are related to obtaining participant informed consent and external researcher authorisation to use data agreement, and for the creation of the participant personal identification code during the pseudonymisation of personal data.

Datasets 4-7. These datasets are analytical datasets produced by the study participants after having obtained both their informed consent and a participant personal identifier code to ensure that no parts of their data can be traced back to the participant during the analysis and publication phases of the research.

As stated above, data collected from participants at UT will be stored and accessed only on UT's server, and data collected from participants at OSU will be stored and accessed only on OSU's server.

(iii) Documentation and Metadata. What documentation and metadata will accompany the data?

As the vast majority of the data is obtained from students on academic writing courses at either UT or OSU, the documentation and metadata accompanying the data are mainly related to the pseudonymisation key identifying string to determine the following:

(1) Type of student artefact. This denotes the type of artefact produced by the participant during the study. Each participant can produce multiple artefacts that include the following:

- Draft text, feedback cover letter, feedback review letter, audio of synchronous writing group meeting online survey, individual interview, focus group meeting, other

(2) Source of artefact. This denotes where and which participant the artefact was produced by and include the following variables:

- Location (UT, OSU, or other)
- Course or Writing Group (Name and year of course)
- Writing Group Number (Number as applicable to the course)
- Participant unique identifier (Pseudonymised number and letter string)

(3) Instructor identifier. For Study 4 only, expert writing assessors (i.e., writing instructors) will be the main participants of the study. Metadata linking this dataset will be formulated using a similar system. as in (1) and (2) above.

Ethics and Legal Compliance

How will you manage any ethical issues?

I have obtained ethical clearance from the Research Ethics Committee of the University of Tartu to obtain data as to the datasets described previously from participants on academic writing courses and/or writing groups for the duration of the research project (see kooskolastus_368T-18). I will be obtaining ethical clearance from the Internal Revenue Board of OSU to obtain data as to the datasets described previously from participants on academic writing courses and/or writing groups for the duration of the research project.

Ethical issues (abridged)

For a comprehensive account for how this research project complies with all ethical and legal issues from an Estonian perspective, please refer to kooskolasus_368T-18 (added as an appendix to this DMP). Although not drafted yet, this research project will also comply with all ethical and legal issues from a US perspective based upon (and also modified) from the content in kooskolasus_368T-18. Below, I present abridged version of the main ethical issues as contained in kooskolasus_368T-18.

(i) Participants

Participating in this research is not harmful to the Participants either physically or mentally. Furthermore, Participants will be informed that their participation in this research is entirely voluntary, that their participation (or non-participation) will have no effect on their course assessment whatsoever, they can withdraw their participation at any stage with no questions asked, and that they are free to examine their data at any time during and after the study. Although participation in certain aspects of the applicable research instrument (surveys, interviews, and/or focus group meetings) may cause a minor inconvenience with regards to the Participant's time, it is envisaged that their participation in such activities may indirectly help support their writing and feedback practices. The guiding principle is that the Participant will never be manipulated for research purposes. This means that any artefacts produced by the Participants of this study will be given the same amount of feedback and the same assessment criteria (if applicable according to the course design) as any other student on the same course not participating in this study (and vice-versa). Before participation, the Participant is given an overview and data management of the postdoctoral research project and the study in question either orally and/or in writing. Participants who do not wish (or are not selected) to participate in the applicable study can give their informed consent for their pseudonymised written artefacts that they will produce as a natural consequence of their participation on the course for use in later studies. The same Participant is only allowed to participate in one main study (e.g., in *Study 1*, but not in *Study 2* and *Study 3*) where they will also give their informed consent that the analysis of their pseudonymised written artefacts can also be used to inform other studies and that secondary analysis on all their supplied data can be conducted for the purposes of Study 7. They will be further informed that only the principal investigator and collaborating researchers will be the only persons to have access to their data (e.g., their written artefacts will not be shown to other Participants in the USA as in *Study 4* where doctoral students and writing instructors assesses the quality of feedback comments). The Participant has the right to interrupt the experiment for any reason and without giving reasons. The Participant also has the right to request the deletion of any of their collected data at any time without any explanation although no changes can be made to already published data. The Participant, after having been informed of all the facts, must give their informed consent to participate through signing an informed consent form with the Principal Investigator or applicable main collaborating researcher (Djuddah Leijen at UT or Susan Lang at OSU).

(ii) Legal Aspects

As the project involves collecting data from human participants from both Estonia and the USA, there are laws governing the safeguarding and transfer of data from Estonia and the USA, and vice-versa. As the general principal, all e-data and their resulting e-documents (i.e., from the pseudonymisation and then analysis of such documents) collected from participants from UT will be stored on UT's servers only, and all e-data and their resulting e-documents (i.e., from the pseudonymisation and then analysis of such documents) collected from participants from OSU will be stored on OSU's servers only. There are also provisions in the informed consent forms at YT whereby the participants give their consent (or do not give their consent) to the transfer of their personal data to the USA (if applicable) as well as that their data is protected in accordance to the GDPR regulations. I will also abide by all the laws and regulations for safeguarding and the transferring of data collected from participants at OSU according to the instructions of the Internal Review Board of OSU.

(iii) Storage of data

- All electronic data created by both the Participants and the Researcher(s) will be collected using an appropriate and University

approved software tool for the applicable *Study* (e.g., LimeSurvey for UT and Qualtrics for OSU), and all electronic data obtained from UT will be stored securely on the UT's server (e.g., Nextcloud) and from OSU on OSU's server (One Drive).

- Only the Principal investigator (Roger Yallop) and internal collaborators from the UT (e.g., Djuddah Leijen, Piia Taremaa, Helen Hint, and Helena Lemendik), will have access to the Participants' original data collected at UT on a needs-to-know basis only, and they will be the only researchers involved in the pseudonymisation process and storage of all intermediate electronic drafts created as a consequence of this pseudonymisation process.
- Similarly, only the PI (Roger Yallop) and main external collaborator (Susan Lang) will have access to the Participants' original data collected at OSU on a needs-to-know basis only, and they will be the only researchers involved in the pseudonymisation process and storage of all intermediate electronic drafts created as a consequence of this pseudonymisation process.
- All intermediate draft electronic data will be password-protected and stored on the appropriate University's (UT or OSU) server for the purposes of their pseudonymisation and analysis, and then this draft data will be deleted permanently after the corresponding data has been fully pseudonymised (i.e., after all intermediate draft electronic documents of the Participants' data required in the pseudonymisation process have no further useful purpose, these draft documents will be deleted).
- All hard data created by both the Participants and the Researcher(s) (mainly Participant Informed Consent and Researcher Authorisation to use Data Forms (if completed in writing), and the Researcher's written notes, and a back-up copy of the data on a password-protected hard drive will be stored in a locked drawer in the Researcher(s) office at UT for data obtained at UT (Lossi 3-306) and at OSU for data obtained at OSU (4132 C Smith Lab, 174 W. 18th St., Columbus, Ohio, USA) for a period of five years after the applicable Study commences.
- All external researchers (e.g., Prof. Susan Lang from the OSU) will only have access to the pseudonymised data collected from UT, and only with the pseudonymised data they need for their tasked research purpose (e.g., cleaning already pseudonymised data, data validation through second coding, etc.) and they will adhere to the Code of Conduct (see subsection below) to ensure Participant data protection;
- To access the data, the external researcher has to sign, either digitally or in writing, an '*Authorisation to use Data Agreement*' with the Principal Investigator from AVOK (Roger Yallop).
- Any pseudonymised data from UT provided to the external researchers must be stored securely according to the data protection principles of the University of Tartu and GDPR (especially Articles 46.1 and 46.2(e); all data will remain on the University's server, and this data will be password-protected and stored thereafter according to the data protection principles of the University of Tartu;
- The Principal Investigator will destroy all the Participant original data (including their consent forms) and the Researcher Authorisation to use Data Forms no later than five years after the start of this research project (i.e., by 1st October 2027 at the latest).
- All Researcher written notes will also be destroyed after they have no further practical use and no later than five years after the start of this research project (i.e., by 1st October 2027 at the latest).

(iv) Safeguards in transferring data from Estonia to the USA

As a general principal, the principal investigator (Roger Yallop) whilst physically located in the USA and any external collaborator regardless of their present geographical location will only access data obtained from UT participants on UT's server. Similarly, the principal investigator (Roger Yallop) whilst physically located in Estonia or elsewhere outside the USA will only access data obtained from OSU participants on OSU's server. As extra precautionary measures to safeguard any possible transfer of EU personal data to the USA, I have devised the following additional safeguards.

Code of Conduct. I have devised a Code of Conduct to ensure appropriate safeguards for the protection of the transfer of personal data from the Participants at the University of Tartu to countries outside the EU in compliance with GDPR Articles 46.1 and 46.2(e) (see notes (i) and (ii) below). To further ensure the protection of data, this Code of Conduct applies to all collaborating researchers who do not work at the University of Tartu (i.e., external collaborators) regardless of whether their institution is based in the EU or outside the EU. The Code of Conduct for external collaborators are as follows:

1. The principal investigator will brief all collaborating researchers about the scope and nature of the research project, and their role within this research project, and how they must store any supplied and generated research data generated in compliance with GDPR Articles 46.1 and 46.2(e) as stipulated in the steps below.
2. There will be no transfer of Participant original data between the University of Tartu and the Host's Institution.
3. All supplied pseudonymised data, the subsequent data editing, and the resulting documents will be password-protected and conducted only on the University of Tartu's server (e.g., Next Cloud).
4. To ensure compliance, the external researcher will be given viewing and editing permissions, but they will be denied permission to download data from the University of Tartu's server.
5. External collaborators will only be supplied with the pseudonymised Participant data on a needs-to-know basis, and their permissions to access this data and the resulting documents within the University of Tartu's server will be withdrawn after completion of their tasks no later than 30th June 2024 (the end date of this project).
6. Any handwritten notes made by the external collaborators must be kept in a locked drawer in their workplace, and destroyed after their practical use no later than 30th June 2024
7. In addition to the safeguards above, the external collaborator will also observe good research practices as according to their own Host Institution's guidelines to good research practices
8. The external collaborator will have to sign an Authorisation to use Data Agreement, stating that they both understand and they will adhere to this Code of Conduct before being supplied with any pseudonymised Participant Personal data.
9. This Authorisation to use EU Agreement will also be counter-signed by the Principal Investigator (Roger Yallop) or the main collaborator from AVOK (Djuddah Leijen).
10. If practically possible, this agreement can be signed digitally by both Parties where it will be password-protected and then stored on the University of Tartu's server.
11. If obtaining a digital signature from both parties is impractical, then the external collaborator will sign the agreement by hand and send the hand-signed agreement to the Principal Investigator or the main collaborator to their following work address:

University of Tartu, College of Foreign languages and Cultures, Lossi 3- 306, 51003 Tartu.

12. On receipt of this agreement, the Principal Investigator or the main collaborator will counter-sign the agreement, and then store this agreement in a locked drawer in their office at the University of Tartu (Lossi 3- 306)
13. All Authorisation to Use EU Agreements will be destroyed no later than five years after the start of the data collection for this project by 1st October 2027 at the latest.

Notes (i) and (ii)

Article 46.1. "In the absence of a decision pursuant to Article 45(3), a controller or processor may transfer personal data to a third country or an international organisation only if the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available.

Article 46.2(e). "The appropriate safeguards referred to in paragraph 1 may be provided for, without requiring any specific authorisation from a supervisory authority, by:

(e) An approved code of conduct pursuant to Article 40 together with binding and enforceable commitments of the controller or processor in the third country to apply the appropriate safeguards, including as regards data subjects' rights."

Extra Safeguards for Principal Investigator. To further safeguard the protection of Participant personal data, when conducting research in the USA (or outside an EU Member State), the Principal Investigator (Roger Yallop) will adhere to steps 2, 3, 4, 6 and 7 of the Code of Practice for external collaborators as outlined in subsection 9.8. above. However, and unlike the external collaborators, Roger Yallop has access to the Participants' original data. He will also pseudonymise the Participants' original data whilst in the USA, As such, the Principal Investigator will also adhere to the following extra precautionary steps in addition to the Code of Conduct that the external collaborators have to observe:

1. When accessing the Participant's personal data outside the EU, the Principal Investigator (Roger Yallop) will be connected to the University of Tartu's VPN.
2. Roger Yallop will use a separate computer to the one he has been supplied with by Ohio State University to ensure further data security, and this computer will be password-protected.
3. Roger Yallop has made provisions in his Postdoctoral Research Plan to conduct regular research visits back to his secondary place of work at the University of Tartu (Lossi 3- 306, Tartu) to collect, pseudonymise, and analysis Participant personal data , and conduct all other required tasks related to his research project whilst on the territory of Estonia.
4. As such Roger has scheduled the following research visits: (i) 08/12/2022 - 15/01/23 (one month); (ii) 23/02/23 - 10/03/23 (two weeks), (iii) 22/05/23 - 13/06/23 (three weeks), (iv) 4/9/2023 - 18/9/23 (two weeks), and (v) 11/12/23 - 14/01/24 (one month).

Thus, the guiding principle to ensure further safeguards for the protection of personal data, Roger will conduct research on the data collected from the Participants from the University of Tartu in Estonia where possible. Conversely, Roger will conduct research on the data collected from the Participants from Ohio State University in the USA where possible.

FAIR data and resources

1. Making data findable

I have obtained ethical clearance from the Human Research Ethics Committee of the University of Tartu to collect data from participants on academic writing courses and/or groups at the University of Tartu (Kooskolastus_36 8T-18) for the duration of this research project. As it has been deemed that the data collected from the human participants is both personal and sensitive, all data collected is only accessible to the principal Investigator (Roger Yallop) and the named collaborating researchers as to the terms and conditions of the ethical clearance. Therefore, this research project will not be participate in FAIR data.

2. Making data openly accessible

Due to the sensitive nature of the data, all collected datasets will only be accessible to the PI and collaborating researchers. No data will be made readily available to researchers from outside this research project.

3. Making data interoperable

As all the datasets are closed, there will be no interoperability for this study.

4. Increase data reuse

All original data will be destroyed no later than five years after the start of this project. Also, as only the PI and project collaborators have access to the datasets, this question is not relevant.

5. Allocation of resources and data security

There will be no open access of data for this project. As such, this question is not relevant.