
Plan Overview

A Data Management Plan created using DMPTuuli

Title: Koulutuksen tarkoitus ammatillisessa opettajankoulutuksessa

Creator: Mika Alhonkoski

Principal Investigator: Mika Alhonkoski

Data Manager: Mika Alhonkoski

Affiliation: Tampere University of Applied Sciences

Template: General Finnish DMP template

Project abstract:

This study examines how conceptions of the purpose of education are constructed within vocational teacher education. The study focuses on the course Designing the Learning Process in the Vocational Teacher Education programme at Tampere University of Applied Sciences. The aim is to explore how teacher students understand the purpose of education and how these understandings are reflected in their pedagogical planning and reflections on learning.

The research draws on Gert Biesta's framework of educational purposes, consisting of qualification, socialisation, and subjectification. The data consist of approximately 250 student-produced learning process workbooks and written course feedback responses. The workbook data include students' descriptions of learning objectives, learning activities, assessment, participation, and a reflective question concerning the most important goal of the learning process from the student perspective. Course feedback data provide additional insights into students' views on learning, teaching, and the significance of the course.

The study employs a theory-informed qualitative research design. Data will be analysed using qualitative content analysis guided by Biesta's conceptual framework. The analysis will examine how different educational purposes are represented in pedagogical planning, how students describe the purpose of education in their feedback, and the extent to which these perspectives align.

The original workbook submissions contain direct identifiers because they are submitted through the Moodle learning environment as part of normal course activities. Only workbooks from students who provide research consent will be included in the study. All direct identifiers will be removed immediately when the research dataset is created, and the analysis will be conducted using anonymised data.

The study will contribute new knowledge about educational purposes in vocational teacher education and provide insights into how future teachers conceptualise the aims of teaching, learning, participation, and student agency. The findings can support the pedagogical development of vocational teacher education.

ID: 32278

Start date: 18-10-2026

End date: 01-01-2027

Last modified: 14-09-2026

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Koulutuksen tarkoitus ammatillisessa opettajankoulutuksessa

1. General description of the data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

The research is based on qualitative textual data collected from students enrolled in the Vocational Teacher Education programme at Tampere University of Applied Sciences.

Two complementary datasets will be used:

1. Learning process workbooks produced by students during the course Designing the Learning Process. The workbooks include descriptions of learning objectives, learning activities, assessment practices, learning environments, student participation, and responses to a reflective question concerning the most important goal of the designed learning process.
2. Written course feedback provided by students at the end of the course. The feedback includes reflections on learning, pedagogical thinking, teaching, and the perceived significance of the course.

The data are generated as part of normal course activities and are subsequently used for research purposes from those students who provide consent for the use of their materials in research.

The original workbook submissions contain direct identifiers because they are submitted through the Moodle learning environment using students' names. When the research dataset is created, all direct identifiers are removed immediately and replaced with anonymised records. Data analysis is conducted only on the anonymised dataset.

No external datasets, secondary datasets, or previously archived datasets will be reused in the study.

File formats

The collected data will primarily consist of:

- Microsoft Word documents (.docx)
- Portable Document Format files (.pdf)
- Moodle exports and administrative files (.xlsx, .csv)
- Anonymised qualitative analysis files (.docx, .xlsx, .txt)

Estimated size of the dataset

The study is expected to include approximately:

- 250 learning process workbooks
- Associated course feedback responses

The complete anonymised research dataset is estimated to be approximately 50-150 MB, depending on the length of workbook submissions and the format of exported files. The dataset is therefore relatively small and can be securely stored within institutional research storage systems.

1.2 How will the consistency and quality of data be controlled?

Data quality will be ensured through a systematic data management and processing procedure. The learning process workbooks and course feedback responses are generated within a single course (Designing the Learning Process) and follow common course instructions and structures, which supports consistency across the dataset.

The original workbook submissions are collected through the Moodle learning environment as part of normal course activities. Because submissions contain students' names and other potential identifiers at the point of collection, only materials from students who have provided consent for research use will be transferred to the research dataset. During this transfer process, all direct identifiers will be removed immediately and the data will be anonymised before analysis.

Data quality will be monitored through the following procedures:

- Verification that only consented materials are included in the research dataset.
- Removal of direct identifiers before analysis.
- Checking files for completeness and readability before importing them into the research dataset.
- Maintaining a clear separation between the original course materials and the anonymised research dataset.

- Creating a structured file naming convention and documentation describing the origin and processing of each dataset.
- Storing all research data in secure institutional storage systems with controlled access.
- Maintaining an audit trail documenting data cleaning, anonymisation, and analysis decisions.

Because the study relies on naturally occurring textual data produced during authentic learning activities rather than questionnaires or measurement instruments, data quality is primarily ensured through careful data preparation, consistent anonymisation procedures, systematic documentation, and transparent qualitative analysis practices.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

The study involves the processing of human-participant data and is therefore subject to applicable data protection and research ethics regulations. Data management will comply with:

- The General Data Protection Regulation (GDPR; EU 2016/679).
- The Finnish Data Protection Act (1050/2018).
- The data protection and information security policies of Tampere University of Applied Sciences (TAMK).
- The ethical principles for research involving human participants issued by the Finnish National Board on Research Integrity (TENK, 2023).

The legal basis for processing personal data is the performance of a scientific research task carried out in the public interest. Participation in the study is voluntary, and only materials from students who have provided consent for research use will be included in the research dataset.

The original learning process workbooks contain direct identifiers because they are submitted through the Moodle learning environment as part of normal course activities. These identifiers may include student names and institutional user information. When the research dataset is created, all direct identifiers will be removed immediately and the data will be anonymised before analysis. The anonymised dataset will be used for all analytical and publication purposes.

Only data necessary for the research objectives will be processed, following the principle of data minimisation. Access to the research data will be restricted to members of the research team, and data will be stored in secure institutional systems with appropriate technical and organisational safeguards.

Research findings will be reported only in aggregated or anonymised form, ensuring that individual students, workplaces, or other identifiable persons cannot be recognised from publications or presentations.

2.2 How will you manage the rights of the data you use, produce and share?

The research data are produced within the context of the Designing the Learning Process course in the Vocational Teacher Education programme at Tampere University of Applied Sciences. Students retain authorship of the texts they produce as part of their coursework. The use of these materials for research purposes is based on voluntary participation and informed consent.

Only materials from students who have explicitly granted permission for research use will be included in the research dataset. The consent process clearly informs participants about the purpose of the study, the use of the data, the handling of personal information, and the possibility that anonymised data may be used in future research.

The original workbook submissions are collected through the Moodle learning environment and therefore initially contain direct identifiers such as student names. When the research dataset is created, all direct identifiers are removed immediately and the material is anonymised before analysis. The anonymised dataset will be used for all research, reporting, publication, and potential data-sharing purposes.

The research team will hold the rights necessary to process, analyse, and publish findings based on the collected research data. Access to the research data will be restricted to authorised members of the research team.

Research results will be published in scientific publications, presentations, and other academic outputs. Any quotations included in publications will be carefully anonymised to prevent the identification of individual participants.

Where legally and ethically feasible, anonymised research data may be retained for future scientific research. Any future use of the data will comply with applicable data protection legislation, research ethics principles, and the conditions communicated to participants during the consent process. No identifiable personal data will be shared outside the research team.

3. Documentation and metadata

3.1 How will you document your data in order to make it findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?

The research data will be documented systematically throughout the project to support transparency, traceability, and potential future reuse of the anonymised dataset. Documentation will be created in English and stored together with the research data within the secure storage environment provided by Tampere University of Applied Sciences.

A project-level README file will be created to describe:

- the purpose and scope of the study,
- the research questions,
- the composition of the dataset,
- data collection procedures,
- consent procedures,
- anonymisation processes,
- file structures and naming conventions,
- data processing and analysis procedures, and
- conditions for future reuse of the anonymised data.

Separate documentation files will also describe the different data sources, including learning process workbooks and course feedback data, their collection context, and any transformations made during data preparation.

A data dictionary will be maintained for any structured datasets (e.g., Excel or CSV files derived from the original materials). This documentation will explain variable names, coding schemes, and classifications used in the analysis.

Version control will be applied to major data-processing stages so that changes to the dataset, anonymisation procedures, and analytical files can be traced over time. An audit trail documenting key research decisions will also be maintained.

The original workbook submissions contain direct identifiers because they are submitted through the Moodle learning environment as part of normal course activities. During the creation of the research dataset, all direct identifiers are removed and the anonymisation procedures are documented in the project documentation. The research dataset used for analysis therefore consists only of anonymised materials.

Data will be stored in widely used and interoperable file formats whenever possible, including:

- DOCX and PDF for textual documents,
- XLSX and CSV for structured datasets and coding files,
- TXT and PDF/A for long-term preservation when appropriate.

If anonymised data are retained for future research use, accompanying documentation will allow other researchers to understand the origin, structure, content, and limitations of the dataset without access to any identifiable participant information. This documentation will support the FAIR principles by improving the findability, accessibility, interoperability, and reusability of the research data.

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will the data be backed up?

The research data will be stored in secure information systems provided by Tampere University of Applied Sciences (TAMK). Access to the data will be restricted to members of the research team through personal user accounts and institutional access controls.

The original learning process workbooks are submitted through the Moodle learning environment as part of normal course activities. These original submissions contain direct identifiers, such as student names. Only materials from students who have provided consent for research use will be transferred to the research dataset. During this transfer process, all direct identifiers will be removed immediately, and the anonymised research dataset will be stored separately from the original course materials.

The anonymised research data, documentation files, analysis files, and project records will be stored in secure institutional storage locations with regular backups provided by TAMK's information technology services. Backup procedures are managed centrally by the institution and are designed to protect against accidental data loss, hardware failures, and other technical incidents.

Only the research team will have access to the working research dataset. Access rights will be reviewed and maintained throughout the project to ensure that only authorised persons can access the data.

No research data will be stored on personal devices, unsecured cloud services, or portable storage media without appropriate encryption and institutional approval. If temporary local copies are required for analysis, they will be stored on password-protected and encrypted institutional devices and deleted after use.

The anonymised dataset and accompanying documentation will be retained according to TAMK's research data management practices and the project's data retention plan.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

The principal investigator, Mika Alhonkoski, will be responsible for overseeing access management throughout the project. Access to the research data will be limited to authorised members of the research team who require access for research purposes.

The original learning process workbooks are submitted through the Moodle learning environment as part of normal course activities and initially contain direct identifiers, such as student names. Access to these original course materials is governed by the access rights and information security procedures of Tampere University of Applied Sciences.

When the research dataset is created, only materials from students who have provided consent for research use will be transferred to the research dataset. Direct identifiers will be removed immediately during this process, and the anonymised research dataset will be stored separately from the original course materials.

Access to the research dataset will be controlled through:

- personal institutional user accounts,
- password-protected systems,
- role-based access permissions,
- secure institutional storage environments,
- restricted access granted only to authorised research team members.

The principal investigator will be responsible for granting, reviewing, and removing access rights when necessary. Access rights will be limited to the minimum level required for conducting the research.

No identifiable research data will be shared outside the research team. If anonymised data are retained for future research use, any future access will be managed in accordance with applicable legislation, research ethics principles, participant consent, and Tampere University of Applied Sciences' data management policies.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?

The original research data cannot be made openly available because the source materials are educational documents produced by students within a university course and the original submissions contain personal identifiers. Although all direct identifiers are removed when the research dataset is created, the qualitative nature of the text data may still contain information that could increase the risk of indirect identification.

For this reason, the full research dataset will not be published as open data.

However, the following materials may be made publicly available:

- metadata describing the study and the dataset,
- the research data management documentation,
- the codebook and analytical framework used in the study,
- descriptions of the data collection and anonymisation procedures,
- aggregated research results,
- carefully selected and fully anonymised quotations included in scientific publications.

The metadata and documentation may be made available through institutional or disciplinary research information systems, repositories, or publication platforms associated with Tampere University of Applied Sciences.

The decision regarding long-term preservation and possible controlled access to anonymised data will be made after completion of the project and after evaluation of data protection, ethical, and legal considerations. If future reuse is considered feasible, access would be restricted to research purposes and subject to appropriate access procedures.

Metadata and research outputs will become publicly available when the study results are published in scientific or professional publications. The dataset itself will remain access restricted unless a later assessment demonstrates that anonymised data can be shared safely and in compliance with participant consent, data protection legislation, and research ethics requirements.

5.2 Where will data with long-term value be preserved, and for how long?

The primary research data consist of qualitative text materials produced by students as part of their studies. The original submissions contain direct identifiers at the point of collection because the learning process workbooks are submitted through the Moodle learning environment using students' names. Direct identifiers will be removed immediately when the research dataset is created, and all analysis will be conducted using anonymised data.

Only the anonymised research dataset and the accompanying documentation will be considered for long-term preservation. The original identifiable course materials will not be archived for long-term research preservation beyond what is required by institutional educational record retention practices.

The anonymised research data, documentation files, coding framework, and metadata will be stored in secure institutional storage maintained by Tampere University of Applied Sciences. At the end of the project, the long-term value of the anonymised dataset will be assessed based on its scientific relevance, reusability, ethical considerations, and data protection requirements.

If the dataset is considered to have long-term research value and can be preserved without compromising participant confidentiality, the anonymised data and related documentation may be retained for future scientific research for at least five years after the completion of the project, and potentially longer in accordance with institutional data management policies and applicable legislation.

Any long-term preservation will include sufficient documentation (README files, metadata, and descriptions of data processing and anonymisation procedures) to ensure that the dataset remains understandable and reusable in the future. Access to preserved data will remain restricted and subject to ethical, legal, and data protection considerations.

6. Data management responsibilities and resources

6.1 Who (for example role, position, and institution) will be responsible for data management?

The overall responsibility for research data management rests with the Principal Investigator, Mika Alhonkoski, PhD, Senior Lecturer, Vocational Teacher Education, Tampere University of Applied Sciences (TAMK).

The Principal Investigator is responsible for:

- ensuring compliance with data protection legislation, ethical guidelines, and institutional policies,
- managing participant information and consent procedures,
- overseeing data collection and anonymisation,
- controlling access rights to the research data,
- ensuring secure storage and backup of the data,
- maintaining research documentation and metadata,
- supervising the retention and disposal of research data, and
- ensuring that research data are managed according to the approved data management plan.

The original learning process workbooks are submitted through the Moodle learning environment and initially contain direct identifiers such as student names. The Principal Investigator is responsible for ensuring that only data from consenting participants are transferred to the research dataset and that all direct identifiers are removed immediately during the creation of the research dataset.

If other researchers participate in the project, they will have access only to the data necessary for their research tasks and will be required to follow the project's data management, information security, and confidentiality procedures.

The secure storage infrastructure, user authentication systems, access control mechanisms, and backup services are maintained by Tampere University of Applied Sciences' information technology services, while responsibility for the research data and its proper management remains with the Principal Investigator.

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

The project requires only modest resources for data management because the dataset consists primarily of qualitative text documents and associated documentation files. No specialised high-capacity storage infrastructure, computing resources, or proprietary software environments are required.

The following resources will be used to support FAIR data management principles:

- Secure institutional storage provided by Tampere University of Applied Sciences (TAMK) for storing research data, documentation, and analysis files.
- Institutional backup services maintained by TAMK's IT services.
- Access-controlled user accounts and authentication systems for secure data management.
- Standard office software for data processing and documentation.
- Widely used and interoperable file formats (e.g., DOCX, PDF, PDF/A, XLSX, CSV, and TXT) to facilitate long-term accessibility and reuse.
- Project documentation, including README files, data descriptions, coding frameworks, and metadata records.
- Researcher time for data anonymisation, documentation, quality control, and preparation of metadata.

The original learning process workbooks contain direct identifiers because they are submitted through the Moodle learning environment as part of normal course activities. Resources are therefore required for anonymisation procedures, including the removal of names and other identifying information before the research dataset is created. All subsequent analysis and preservation activities will be based on the anonymised dataset.

Because the dataset is relatively small (estimated at less than 150 MB), no additional funding, external storage services, or dedicated preservation infrastructure are anticipated. Existing institutional resources are sufficient to ensure secure storage, documentation, long-term preservation, and potential future reuse of the anonymised data in accordance with FAIR principles and applicable legal and ethical requirements.

Planned Research Outputs

Journal article - "Koulutuksen tarkoitus ammatillisessa opettajankoulutuksessa"

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Koulutuksen tarkoitus ammatillisessa opettajankoul... ...		Journal article	Unspecified	Open	None specified		None specified	None specified	No	No