Plan Overview

A Data Management Plan created using DMPTuuli

Title: Exploring the Longitudinal Impact of Generative AI Feedback on Student Engagement and Learning Outcomes in STEM Education

Creator:john boakye asibuo

Principal Investigator: Asibuo Boakye John

Contributor: Mohammed Saqr

Affiliation: University of Eastern Finland

ORCID iD: 0009-0005-0069-9004

Template: General Finnish DMP template

Project abstract:

Abstract
This study investigates the longitudinal effects of Generative AI (GenAI)-driven feedback on student engagement and learning outcomes in STEM education, focusing on teacher training colleges in Ghana’s Upper West Region. GenAI tools offer personalized, real-time feedback that promises to enhance iterative learning and reduce instructor workload; however, their sustained impact on student engagement–encompassing behavioral, emotional, and cognitive dimensions–remains underexplored. Employing a two-phase mixed-methods research design, Study 1 establishes baseline engagement metrics and develops tailored GenAI feedback methods, while Study 2 delivers weekly personalized feedback and tracks its influence on engagement trends and academic performance over time. Both quantitative (LMS activity logs, surveys, test scores) and qualitative (interviews) data will be analyzed to assess engagement, outcomes, and context-specific adoption factors. The research also addresses critical ethical considerations, including data privacy, algorithmic bias, and accessibility. By focusing on longitudinal engagement rather than short-term effects, this work provides novel insights into the capacity of GenAI interventions to transform learning in underserved STEM environments, informs equitable policy decisions, and suggests strategies for the ethical and effective deployment of AI tools in education.

ID: 29422

Start date: 25-05-2025

End date: 21-12-2029

Last modified: 02-11-2025

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

Exploring the Longitudinal Impact of Generative AI Feedback on Student Engagement and Learning Outcomes in STEM Education

---

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.

- **Types of Data Collected/Produced:**
    1. **Quantitative Data:**

i. Learning management system (LMS) activity logs (e.g., log-ins, assignment submissions, forum participation)
ii. Weekly survey responses (e.g., Likert-scale questions on motivation and engagement)
iii. Problem-solving task scores (weekly, focused on cognitive engagement)

    1. **Qualitative Data:**

i. Semi-structured interview transcripts from students and instructors about their experience with Generative AI-based feedback

- **File Formats:**
    1. CSV, Excel, or other spreadsheet formats (for logs and survey results)
    2. Text, RTF, or DOCX (for transcribed interviews)
    3. Audio files (if interviews are recorded)
    4. Statistical software files (e.g., SPSS, R) for data analysis
- **Data Reuse:**
    1. No direct reuse of existing datasets is indicated that the data will be newly collected as part of the study.
- **Estimated Data Size:**
    1. Approximately 5–10 GB for all raw data, transcripts, and backup files across the project duration.

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

Consistency and quality of data will be controlled through several measures :

a. Standardized protocols: Surveys, cognitive tasks, and LMS activity logs will be collected using standardized procedures to ensure comparability across participants and over time.

b. Pilot testing: Instruments and surveys will be piloted with a subset of participants to identify and correct potential issues before full-scale data collection.

c. Double-checking and audits: Data entries (especially from manual processes like interview transcription) will be cross-checked by a second coder or auditor to minimize errors.

d. Use of reliable digital platforms: Data will be collected and managed using reliable, institutionally-supported platforms (such as Moodle, Google Classroom, or secure survey software) to reduce procedural variability.

e. Regular review and validation: The research team will regularly review collected data for inconsistencies, missing values, and anomalies, implementing corrective measures if needed.

f. Clear coding schemes and documentation: All variables, scales, and coding decisions will be clearly documented to maintain consistency in analysis.

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.)

Legal issues related to data management in this research include compliance with GDPR and similar data protection laws. This involves ensuring data privacy, particularly when handling personal data from LMS logs and surveys. Participants must be informed and provide consent for their data usage, with explanations tailored to local cultural contexts. Data must be anonymized to protect individual identities. Another key legal aspect is the need to test algorithms for fairness and minimize algorithmic bias across diverse student demographics. These measures align with both ethical standards and legislative requirements governing data processing, especially for research involving personal and sensitive information

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

The rights to the data collected, produced, and shared in this research will be managed in the following ways:

a. Informed Consent: All participants will be fully informed about the data usage policies and must provide explicit consent, tailored to local cultural contexts, before data collection.

b. Data Ownership: Data rights will be held collectively by the research team and the host institutions. Any use, publication, or sharing of the data beyond the immediate team will require appropriate internal approvals.

c. Anonymization: Only anonymized and aggregated data will be shared or published, ensuring participant identities are protected.

d. Data Sharing: Sharing of data outside the project will be limited to aggregated results or excerpts, and only with explicit permission from participants and under the conditions specified by consent agreements and institutional policies.

e. Algorithmic Fairness: Special attention will be given to minimize algorithmic bias when using or sharing algorithm-driven feedback data, ensuring fair representation of all demographic groups.

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?

Data will be documented for findability, accessibility, interoperability, and reusability as follows:

a. README file: Each dataset will include a comprehensive README file describing the data collection process, the purpose of each file, variable definitions, coding schemes, and any preprocessing steps.

b. Metadata standards: Metadata (such as variable names, units, date/time stamps, and anonymization status) will be included within datasets and files, following either simple Dublin Core principles or local institutional guidelines.

c. Consistent labeling: Interview transcripts will be labeled with pseudonyms, dates, and clear file identifiers. Survey and log files will use standardized variable names and coding.

d. Documentation: All surveys, coding schemes, and protocols will be clearly documented and included as separate supplementary files to facilitate reuse and understanding by other researchers.

e. Interoperability: Data will be saved in widely accepted formats (e.g., CSV, XLSX, TXT) to ensure compatibility with various software. Documentation will detail formats and structures.

f. Accessibility: All documentation and data files will be kept in an organized folder hierarchy with logical naming conventions, supporting easy access and understanding for all team members and potential future users.

4. Storage and backup during the research project

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

Data will be stored on secure, password-protected cloud servers provided by the host institution or collaborating universities. Regular automatic backups will be performed, with at least one encrypted offsite backup to ensure data safety and continuity in case of hardware failure or other incidents. Only project personnel with authorized access will be able to access and manage the data, and institutional IT security protocols will be strictly followed throughout the project.

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

The Principal Investigator (PI) will be responsible for controlling access to the data. Secure access will be managed through institutionally supported IT systems, with strong password protection and, if available, two-factor authentication. Only authorized members of the research team will be granted access; permissions will be reviewed regularly to ensure compliance with ethical and legal requirements. Physical and digital security measures (such as encrypted backups and access logs) will

be implemented to maintain data integrity throughout the project

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?

Parts of the data that can be made openly available or published will include:

a. Aggregated, fully anonymized datasets (such as survey results, engagement metrics, and coded interview themes) that do not contain any personal identifiers.

b. Metadata describing the datasets will also be made available so that others can interpret and reuse the data.

c. The data and metadata will be deposited in a recognized public repository (such as a national or institutional data archive) after the research is completed and results published.

d. Sensitive raw data (such as full interview transcripts or logs containing personal information) will not be published but may be shared upon request with appropriate ethical approvals and only in anonymized form.

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

Data with long-term value will be preserved in a recognized national or institutional data repository, such as the Finnish Social Science Data Archive or a similar accredited archive. The retention period for such data will be a minimum of 5–10 years, or longer if required by funder or institutional policy. This ensures future accessibility for verification, reuse, and further research, while respecting ethical guidelines on data retention and participant privacy

6. Data management responsibilities and resources

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

The Principal Investigator (PI) at the host institution (such as NJA College of Education Wa or McCoy College of Education Nadowli) will be responsible for data management. The PI will be supported by an institutional data manager or relevant IT personnel to ensure all ethical, legal, and technical requirements for data handling, storage, and sharing are met

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)?

1) Technical Infrastructure
a) Secure cloud storage with encryption and backup systems
b) LMS platforms (Moodle, Google Classroom) with data export capabilities
c) Survey and assessment tools for data collection
2) Metadata & Documentation
a) DOIs and metadata schemas for dataset identification
b) Data dictionaries and codebooks
c) Open format conversion tools (CSV, JSON, XML)
3) Human Resources
a) Data manager trained in FAIR principles
b) Technical support staff for LMS integration
c) Ethics compliance officer
d) Transcription services for qualitative data
4) Privacy & Security
a) Data anonymization software
b) Encrypted transfer protocols
c) Consent management systems
d) Legal compliance review
5) Access & Sharing
a) Open access repositories (Zenodo, OSF, institutional)
b) Data sharing agreements and access protocols
c) Version control systems
d) Published analysis scripts and documentation
6) Financial
a) Cloud storage subscriptions
b) Repository hosting and DOI fees
c) Software licenses
d) Personnel and training costs
7) Context-Specific (Ghana)
a) Reliable internet connectivity
b) Local technical support
c) Capacity building programs
d) Partnerships with African research repositories
8) These resources ensure the study's longitudinal data can be properly stored, documented, shared, and reused while maintaining ethical standards and addressing infrastructural challenges in the Upper West Region.