

# Code & Data

## Experiences of Learning to Code

Joe Marsh Rossney

August 29, 2025

This page describes and locates the various code and data artifacts produced during the project.

### Available code & data

- JSON file<sup>1</sup> with which the Jisc survey can be reconstructed.
- Survey data (to do)
- Code for analysing and visualising survey data (to do)
- Python tool<sup>2</sup> used to format the interview transcripts.

### Why not publish the interview transcripts?

Where research involves real human lives, there can be a tension between the transparency and reproducibility goals of the researcher, and the obligation to take every reasonable precaution to protect the privacy and security of the people involved.

We intended to strike a balance by publishing a collection of the most relevant sections of interviews, after making certain to redact potentially identifiable or irrelevant information, but not the interviews in their entirety.

However, upon reflection we have decided not to publish this dataset, for the following reasons:

---

<sup>1</sup>[https://github.com/ExpLrnCode-2024/jisc-surveys/tree/main/jisc\\_export/survey.json](https://github.com/ExpLrnCode-2024/jisc-surveys/tree/main/jisc_export/survey.json)

<sup>2</sup><https://github.com/ExpLrnCode-2024/teams-transcript-formatter>

## Ambiguity in the informed consent form

The informed consent form<sup>3</sup>, signed by all participants prior to their interview, gives permission to publish “*sections of the interview*” in “*research outputs and websites*”.

There are two main issues with this.

First, students might reasonably assume that “research outputs” means communication documents such as articles and websites; we should have explicitly included “dataset” in the list of potential outputs if that was our intention.

Second, when the dataset is basically “the interview, minus sensitive or irrelevant information and false starts”, although this is *technically* “sections of the interview”, it does not feel reasonable to say that this is covered by the informed consent form.

We could heavily cut down on the number and length of interview sections included in the dataset, but the value of this ‘dataset’ as a research object falls off very quickly as the context surrounding each section of the interview is stripped away.

## Automated analyses

Our motivation for publishing the collection of interview sections was the hope that other researchers might perform their own analysis, uncovering any aspects we missed or insights that pertain to a different research question than ours.

However, the even in the last year the research landscape has shifted in such a way that substituting qualitative analysis for Large-Language-Model summaries is considered not only acceptable but innovative, at least by some.

I do not share this view, but this is largely irrelevant since the participants did not consent to their data being processed in this way.

We suspect that a dataset such as this, made open-access in a convenient plain text form, would be attractive to individuals looking to do automated analysis — probably far more so than it would be for researchers who prefer traditional methods.

## Risk of identifiability

We made a significant effort to redact all sensitive or potentially identifiable information from the interview transcripts prior to carrying out our main analysis.

We are confident that a human would find it extremely difficult to identify an individual based on reading the redacted transcripts.

---

<sup>3</sup>[jisc-surveys/participant\\_info.html](https://jisc-surveys/participant_info.html)

However, publishing the dataset has the unfortunate side effect of making it available to data ingestion engines. This increases the risk that an individual may be identified through correlations between the information in the transcript and other information online.

## Authors

During the relevant time period (2024), all authors were affiliated with the School of Physics & Astronomy at the University of Edinburgh. **Joe Marsh Rossney** had recently completed a PhD in theoretical physics, during which time they were a teaching assistant on several different programming courses. **Sarah Hogarth** had recently completed a Bachelors degree in physics, where their dissertation focused on the impact of Generative AI on physics education. **Polux Gabriel Garcia Elizonda** was a Master's student in physics, having also completed a dissertation on Generative AI in physics education. **Ross Galloway** was a Senior Lecturer and leader of the Physics Education Research Group. **Britton Smith** was a Reader in the Institute for Astronomy and Course Organiser for an introductory Python course taken by physics undergraduates.

## Author contributions

CRedit: **JMR**: Conceptualisation (lead), Data curation (lead), Formal analysis (equal), Funding acquisition (lead), Investigation (lead), Methodology, Project administration (equal), Software, Supervision (of SH & PGGE), Writing - original draft. **SH**: Data curation (supporting), Formal analysis (equal), Investigation (supporting). **PGGE**: Data curation (supporting), Formal analysis (supporting), Investigation (supporting). **RG**: Conceptualisation (supporting), Funding acquisition (supporting), Project administration (equal), Supervision (of JMR), Writing - review & editing. **BS**: Conceptualisation (supporting), Funding acquisition (supporting).

## Acknowledgements

The authors would like to thank **Kristel Torokoff** for playing an instrumental role in securing financial support for this project via the School of Physics and Astronomy. We would also like to thank **Kristel Torokoff** and **Joe Zuntz** for conversations that helped to shape this project.

## Financial support

We gratefully acknowledge that funding for this Principle's Teaching Award Scholarship (PTAS) project was provided by the University of Edinburgh Development Trust.

**JMR** was directly supported by both PTAS and the School of Physics & Astronomy at the University of Edinburgh. **SH** was supported by PTAS. **PGGE** was supported by the School of Physics & Astronomy through the Career Development Summer Scholarship programme.

## **Correspondence**

- [joemar@ceh.ac.uk](mailto:joemar@ceh.ac.uk) for enquiries related to the project, website, code and data.