Project Title: An investigation into the way Jupyter Notebooks enhance learning and teaching on TM351

Key words: Jupyter notebook, iPython, TM351, lab book, code, pandas, python, visualisation, study resources, study location, integrating theory and practice

Authors: Sharon Dawes and Chris Thomson

Date of Report: 14 June 2022

Key staff associated with the project: Stephen Rice and Stephen Bowles

Contact points: sharon.dawes@open.ac.uk and chris.thomson@open.ac.uk

Executive Summary

TM351 Data Management and Analysis has used Jupyter (2019) notebooks for practical work from its inception in 2016 and this project investigated how the notebooks supported students' learning.

Jupyter is a web-based tool that enables students to embed program code and associated discussion in a notebook in the style of a lab book so that they can explain what they were doing and why. Teaching materials include notebooks in the same style, to explain a concept and then show its practical application such that the example can be run. In this way students can return to a project at a later stage and if their notes were sufficient, they or another researcher can repeat the work and study how valid their assumptions were as well as whether their conclusions in a report were justified.

Students completed an online questionnaire after studying the module and this was followed by an in-depth interview to elicit more information about their responses.

Survey respondents reported that they were able to successfully integrate their understanding of the module materials between the VLE (Virtual Learning Environment) website and Jupyter notebooks. Whilst there were some occasional issues in setting up the notebooks most used them without problems, but there were some who found them hard to access, and hard to search. The use of notebooks limited some students' ability to study when and where they wanted to as a full computer is required, and there were concerns about if the notebooks should be used to teach theory and a well as the practical work. This suggests that notebooks may not be suitable for all learning needs or preferences.

Most of the students interviewed studied the module by using the notebooks to practice techniques they had learned about in previous session on theory. While for some there had been problems installing or running module software, the help available to them had enabled them to resolve problems and students had enjoyed the module, albeit finding the workload very heavy.

Aims and scope of the project

The project investigated how Jupyter notebooks supported students' learning on TM351 Data Management and Analysis. It examined how students divided their time between theoretical reading on the TM351 module website and practical work using notebooks, both as a part of their learning, and then practice for their final report. It explored to what extent their use of the notebooks reinforced their learning or extended their knowledge and whether there was any relationship between where they were currently studying and what computing devices they had available when the student worked at home, at work or on the go that might have affected their ability to undertake practical work using notebooks as opposed to theoretical reading using other provided materials.

The research questions the project set out to answer were:

- Does the use of notebooks effectively support students in their studies of the module learning outcomes?
- Do notebooks allow the students to integrate theory and practice?
- Do students have technological problems in using notebooks?

- Does the use of notebooks dictate when and where students are able to study?
- Is the use of notebooks restricted by accessibility constraints?

Activities

Three cohorts of students from three consecutive academic years (2017J: 420, 2018J: 372, 2019J: 365) were surveyed. Cohorts 2017J and 2018J had completed the module, cohort 2019J were midway through the module. Following university procedures to avoid over surveying students, 301 students (who had completed or were completing the module) were randomly selected to be invited to complete the survey. The survey was administered through Online surveys (JISC 2022). There was a total of 55 responses (2017J: 19, 2018J: 14, 2019J: 22). Of these, 29 respondents opted in to be interviewed, and we interviewed 10 of them.

As this was novel research, we did not find any existing questions on which to base our research. The questions were framed based on our experience of supporting the students We felt that the students' approach and experience to study would be more significant than descriptive demographics, so we did not capture them separately given the anticipated number of responses. A full list of the questions is in the appendix of the detailed analysis of the survey (Thomson and Dawes, 2022).

Questions were selected to identify when the student studied the module, familiarity with the technology, problems they had, and how they structured their study. We asked what might improve the teaching of the module in general, as well as suggesting that the taught material and notebook exercises could be combined. The full set of questions are included in the appendix.

We used a mixture of closed demographic questions, Likert scales for agreement and open questions. 45% of respondents included some comments to open questions.

Single-item Likert measurements were used. Whilst less accurate than a multiple item scale, due to the relatively small population, and expected response, we wanted to minimise survey length to have some data. We intended to uncover additional dimensions via the interviews. We selected a commonly used 5-point response option for agreement [Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree].

For free text answers we followed the process of thematic coding as described by Braun and Clarke (2006). As there were relevantly few responses (55) and in each question most responses were 1 to 2 sentences in length, the coding applied was simplistic in nature. The goal was to identify the unique explanations (themes) in each answer in order to get a feel for how many students had similar experiences.

We conducted interviews with 10 students by using Skype for Business and analysed the transcriptions of these.

We asked students:

- how they studied in a typical session and how they integrated their learning on the practical parts of the module with the theoretical parts on the OU website

- how they found information from previously studied topics on the module and how the notebooks enabled them to learn new techniques.
- how they used the notebooks in preparing assignment questions and to what extent the notebooks supported their report writing.

We also explored whether the students' places of study or computer equipment had any bearing on what they were telling us and whether the module would benefit from any additional resources.

Finally, we asked them whether they found the practical or theoretical work easier to study and whether there were any aspects of the module that led to confidence issues.

Findings

1. Summary of Survey

The survey focused on students' reflections on how they could study the module, so it was not a direct measure of how effective it was at supporting learning outcomes. However indirectly, students reported that they felt supported and made use of the tools provided in the anticipated way. In particular, students typically turned to the module website and VLE first when answering questions and then the notebooks, showing good integration of theory and practice and suggesting that they were effectively supported.

Students reported that they followed the expected pathway of starting with studying the theory on the VLE website, then notebooks, followed by computer marked assessments and tutor marked assessments. However, this was not a straightforward linear pathway, with students typically using both the website and notebooks in the same study session. Furthermore, students attempted a large proportion of the exercises provided in the notebooks, suggesting that a deep level of learning was taking place.

Students reported that the software was mostly installed without problems, although some did need additional assistance. However not all features of the notebooks were apparent to the students, and they struggled to find material within the notebooks on demand as they were hard to search.

Whilst the majority of students spent most of their study time on a desktop or laptop computer, this was not always the case. Students studied in a variety of locations and with a variety of devices. It is likely that the fact that notebooks were only available on a full computer skewed the results here. Comments from students highlighted their desire for more diverse study materials, including those which are printed, and how these assist learning in more diverse situations.

Within this sample there were a majority who identified no significant accessibility constraints but some who did. The comments provided focused on how easy the teaching material was to find and how it was presented, highlighting the need for printed material. As with any teaching individual needs will need to be assessed and supported appropriately when using Jupyter notebooks.

More significantly, opinion was divided on whether the theoretical part of the teaching should be included in Jupyter notebooks. One way to interpret this is that students valued the diversity of teaching materials where the best medium was used for each element.

2. Summary of Interviews

Students' experience on the module differed depending on how much previous experience they had and those with insufficient Python experience found the learning curve very steep at the start and would have benefited from more consolidation resources to help with this and basic Pandas and SQL. The more confident students omitted practical work where they already understood the material or perhaps did not need it to undertake an assignment.

Students study sessions began with theoretical work for a topic after which they carried out the practical activities, returning to theory if there were aspects they had not understood. If they were studying several modules, they might do several topics on each module before rotating their study, so this meant they had to revise work next time they came back to TM351.

There were software installation problems for some students, and these were mostly overcome with a provided download, but one student missed several week's work as a result. A common problem was in the length of the notebooks, in terms of amount of code, especially for the EMA, which sometimes made the software crash, having taken many hours to run. Other issues included the number of tabs that were open at once, swallowing up resources.

Theoretical work was undertaken in a variety of places and on different devices or, for 1 student, using printed copies, while practical work was usually undertaken at home on the student's own PC. There was a split of Mac and Windows users as well as one Linux user. Eight out of ten students thought that the practical work was easier.

Students found that the integration of explanations of code snippets and explanations worked really well in helping them to understand concepts but were divided as to whether they would have liked more text to be included here. The fact that you could build on the explorations in the notebooks to answer deeper questions was appreciated by some. They were provided with skeletons for the assignment questions, and they found it useful to copy appropriate code from other notebooks into these. The notebooks were seen as a really useful learning tool but sometimes examples were a little narrow in scope, such as concentrating on Milton Keynes for maps.

One problem with notebooks was the difficulty in finding topics to refer back to. Numbering of code cells is by design not persistent, so it was impossible to reference them in notes. The naming of notebooks helped to some extent with this but otherwise students mostly relied on the list of notebooks related to theoretical parts of the module. A useful resource written by a student consisting of summaries of each of the different formulae and codes used by the various notebooks was also mentioned.

Only 50% of the students took notes and they used Word documents, handwritten notes, and Evernote, a Jupyter notebook or mind maps for this. 60% of students attended several or many tutorials and 40% used either forums or their tutor for help.

The final assignment on TM351 is a report, with supporting evidence being in notebooks and the structure of these did not always match what was required for the report, however the chronological order of the cells in the notebook was appreciated as important. One student commented that some guidance on how to make the notebooks look good for reports and how to

create a notebook that pulled in other notebooks easily and ran all joined up and connected would have been helpful.

The comments students made about what could be improved for the future were very few. They would have liked an additional resource summarising keyboard shortcuts and a way of indexing cells in notebooks. More assignments would have been welcome for one student or for another, smaller tests and quizzes in addition that would help embed the learning.

Overall, the students enjoyed the module, and they especially liked the fact that the notebooks provided immediate results without compilation of code and the visualisations they generated.

Impact

We presented the results and analysis to the TM351 and M269 module teams. The project demonstrated that TM351's variety of resource types were benefitting students well.

Our analysis highlighted a few key issues that the module teams were interested in. The TM351 module team felt that being able to search through a whole folder of Jupyter notebooks was a skill that level 3 students should already have and they highlighted that in fact this hadn't been taught at lower levels. They felt that this was a gap that should be considered by Board of Studies as part of the level 1 refresh.

As a prerequisite for studying TM351, students must have studied Python to the level of M269. While this knowledge was assumed by the course materials, some students who hadn't consolidated their knowledge well enough found that there was a steep learning curve at the beginning. We recommended that the current introduction to Python material be retained in some form in the module materials to bridge any gap the students might have even if such a gap wasn't desirable.

Students who have not studied SQL at level 2 would be at a disadvantage if the module didn't teach skills that should have been acquired earlier in the students' degree studies. The module team recommended that this should be considered by the Board of Studies, as there was a concern that module content would not reach QAA level 3 if too much introductory material was included in level 3 modules.

As students reported that their confidence with notebooks was important in using the examples the teams felt that there was a need to standardise tools across the curriculum. Therefore we will recommend to the Board of Studies that the tools used and staff familiarity is investigated with a view to selecting a small number of standard tools to be used in all computing modules. This would make it easier for students to focus on the academic value of courses not the tools used. We have a concern that some students may value a variety of tools for employability reasons and this should also be considered.

It should be noted that problems reported by students in installing software have been largely eliminated by the provision of the software on Open University servers in more recent presentations. This followed preliminary feedback that we made to the teams.

Next steps

Since we completed this research M269 in 2021 started to teach all module material within Jupyter notebooks. As our findings suggested that students preferred a variety of ways to access the material this presents an opportunity to ask similar questions to see if our findings hold in this situation. M269 has also produced the module materials in an intermediate format so they can more easily create PDF and notebook versions of the module materials. This gives further opportunity to explore the use of formats and may also allow issues such as searching the notebooks to be overcome.

List of deliverables

eSTEeM Final Report and appendices:

2022 An investigation into the way Jupyter Notebooks enhance learning and teaching on TM351

2022 TM351 Jupyter notebooks project - analysis of interviews

2022 Detailed analysis of a Survey on student experiences with Jupyter Notebooks on TM351

Horizons in STEM Conference:

2021 An investigation of how Jupyter Notebooks enhance students' learning of data management - project update

eSTEeM annual conference presentation:

2020 An investigation into the way Jupyter Notebooks enhance learning and teaching on TM351

eSTEeM Project Poster

2019 An investigation into the way Jupyter Notebooks enhance learning and teaching on TM351

References

Dawes S.R., Thomson C.D. (2019) "An investigation into the way Jupyter Notebooks enhance learning and teaching on TM351". Available at: http://www.open.ac.uk/about/teaching to and to learning/esteem/sites/www.open.ac.uk.about.teaching to and to learning.esteem/files/files/2019 to 10 to Dawes to Thomson to poster.pdf

Dawes S.R., Thomson C.D., Rice, S. & Bowles, S (2020) "How Jupyter Notebooks enhance learning and teaching on TM351", presentation at *eSTEeM conference 2020*. Available at: http://www.open.ac.uk/about/teaching to and to learning/esteem/sites/www.open.ac.uk.about.teaching to and to learning.esteem/files/files/2020 to 04 to 29 to 30 to Parallel to Session to D to Sharon to Dawes to et to al.pptx

Dawes S.R., Thomson C.D., Rice, S. & Bowles, S (2021) "An investigation of how Jupyter Notebooks enhance students' learning of data management and analysis", *Horizons in STEM 2021 Conference Proceedings*, pp.30-31. Available at: https://www.open.ac.uk/about/teaching-and-learning/esteem/sites/www.open.ac.uk.about.teaching-and-learning.esteem/files/files/2021-06-Horizons-in-STEM-HE-Conference-Proceedings-FINAL.pdf

Dawes S.R., Thomson C.D., Rice, S. & Bowles, S (2022) "TM351 Jupyter notebooks project - analysis of interviews", in press.

JISC (2022), Online Surveys, Available at: https://www.onlinesurveys.ac.uk/

Thomson, C.D, Dawes S.R, (2022) " Survey on student experiences with Jupyter Notebooks on TM351: Detailed analysis ". In press

University approval processes

- SRPP/SSPP – Approval from the Student Research Project Panel/Staff Survey Project Panel was obtained according to the Open University's code of practice and procedures before embarking on this project. Application number 2019/155

- Ethics approval was granted by the Open University (HREC/3456/Dawes)
- Data Protection Impact Assessment/Compliance Check – A Data Protection Impact Assessment/Compliance Check was obtained according to the Open University's code of practice and procedures before embarking on this project. Data Protection registration number 28-04-26