National Data Skills learning and pilots

Final report

For the Office for Students

By the Careers Research & Advisory Centre (CRAC)

4 August 2022

Contact: Robin Mellors-Bourne
Director, Research and Intelligence, CRAC

Careers Research & Advisory Centre (CRAC) Ltd
22 Signet Court
Cambridge CB5 8LA

robin.mellors-bourne@crac.org.uk

www.crac.org.uk

Contents

Executive summary

Recognising the significance of data skills for the future of the UK economy, as highlighted in the UK Government's National Data Strategy in 2021-22, the Department for Digital, Culture, Media & Sport (DCMS) provided funding through the Office for Students (OfS) to explore how universities were supporting undergraduate students to gain data skills. This programme of work aimed to showcase and demonstrate some of the approaches being utilised in teaching foundational data skills to students pursuing courses without a significant data science element ('non-cognate' students) and attempted to gain any insights into which approaches work best.

Funding was allocated as part of the Postgraduate Conversion Course programme to support seven UK universities to showcase their current provision of this type, while the Careers Research & Advisory Centre (CRAC) was commissioned to assist them in considering the effectiveness of their provision and identification of good practice. The seven 'pilot projects' received funding from October 2021 to March 2022 and concluded delivery of their funded activity in May 2022.

CRAC was also commissioned to undertake a wider evidence review to identify how providers within and beyond the UK teach data skills to non-cognate students (i.e. those on courses without a significant data science element). This review took the form of an initial exploratory literature search, together with a web search exercise to identify provision of data skills modules and other relevant learning by universities and/or third parties.

The findings of this international evidence review, as well as the emerging findings from the pilot projects undertaken at seven UK universities, are presented here.

Data skills

There continues to be a lack of clarity about the scope and language of 'data skills' (as opposed to broader 'data literacy', for example) and competencies, and the extent to which these should be considered 'foundational' in higher education (HE) institutions. Some of the pilot projects reported misalignment between what the skills industry states it seeks and what students are taught (which can be focused on what is needed to undertake a project or dissertation within their programme).

Pilot projects also reported some students being unable to acknowledge the value of data skills outside their degree, such as within their own potential career, suggesting scope for more work contextualising this type of learning and demonstrating its relevance across a wide range of careers. There was also some evidence of 'statistics anxiety' (and similar anxieties) amongst non-cognate students in particular, and that teaching such students can be pedagogically challenging where there are wide variations in levels of data skills.

These issues could to some extent reflect the relative immaturity of data skills as a 'discipline' and also their wide applicability across many non-cognate subjects. In contrast, there is a longer history of teaching statistics in subjects such as psychology, for example, from which it may be possible to learn.

Approaches and effectiveness

A wide range of approaches to data skills teaching (or learning) was seen, ranging from embedded content or mandatory modules within subject curricula taught by the subject team or through service teaching, right through to encouragement to undertake self-study using third party resources. There was some evidence that relatively few undergraduates take up data skills programmes or learning where offered outside the subject curriculum. There seems as yet to be little published evidence about the effectiveness of these approaches, specifically in relation to data skills.

A number of the pilot projects found that modules embedded within the subject curriculum, featuring examples located within the subject, were more popular with students than standalone modules or content provided through service teaching. ‘Real world’ applications and examples, relevant to the non-cognate subject context, were the most appealing. Tying data skills learning into wider employability development could also be effective.

Pilot projects reported perceptions that the earlier in a course that data skills are taught, the better – giving time to embed these skills and demonstrate their importance well ahead of when they need to use them in final projects/dissertations. There was some anecdotal evidence that an optional module introducing data skills early in a course could increase attainment in subsequent modules.

Implications

A lack of evaluative evidence about the effectiveness and impact of different approaches to teaching data skills remains. However, there is emerging evidence that data skills learning embedded in a subject curriculum, and taught by the subject team, may work best for non-cognate undergraduates. This has implications for the upskilling of subject teaching staff (and could be more resource-intensive than a service teaching model), and possibly merits a national upskilling programme for HE staff.

Although the pilot projects reported here were modest, some have started to scale up their activities and translate emerging findings into other contexts, using other funding sources, while others have embedded particular models into other courses as a result of the pilot.

Other observations

There was a relatively high level of interest from HE providers to do more work in the area of addressing foundational data skills learning. Several pilot project teams indicated they would continue to develop and evaluate their approaches. In practice, the extent of evaluation possible within a short funding period was limited, and not all learners were prepared to engage in such research in addition to the requirements of their course.

Recommendations

1. DCMS should work with key stakeholders to develop and agree appropriate terminology and definitions for data skills and data literacy, including the extent to which these should be considered ‘foundational’ in the context of HE provision.

2. The sector should undertake further strategic consultation with industry to establish the types of data skills that employers are seeking from graduates, to ensure that curricula and other data skills learning is as fit-for-purpose as realistically possible.

3. The sector should consider a focused programme of work to develop a range of materials demonstrating the value of data skills to those working in a wide range of early careers, so that more students appreciate their potential future value.

4. Providers need to undertake longer-term, substantive evaluation of different approaches to data skills teaching, building on the pilot projects in this study, in order to obtain a more robust evidence base about the effectiveness of approaches.

5. The sector should consider the potential value of a national programme to upskill non-cognate HE teaching staff to build their confidence in delivering data skills teaching, highlight best practice in teaching students with mixed abilities and prior experiences, and share practice and resources.

1 Introduction and context

1.1 The importance of data skills

The National Data Strategy (NDS), published in September 2020, set out a vision to harness the power of responsible data use to boost productivity, create new businesses and jobs, improve public services, support a fairer society, and drive scientific discovery, positioning the UK as the forerunner of the next wave of innovation.[1] It is a framework for the action required to maximise the power of data across the UK. One of its four key pillars is data skills – with a stated desire that the right skills are developed through the education system and that people can continue to develop the data skills they need throughout their lives.

The need for data skills continues to grow across the economy. The Royal Society has reported that demand for specialist data skills has more than tripled since 2013,[2] while analysis of over nine million online job adverts predicted that data analysis skills will be the fastest growing digital skills cluster over the next five years. Recent research has found that UK companies were seeking to fill up to 234,000 roles requiring hard or technical data skills and almost half of businesses have been recruiting for data-related roles.[3] Research from Forrester in 2021 has found that recruiters rank data literacy as the skill highest in demand for entry-level candidates.[4]

Those and other reports suggest exponential growth in the demand for advanced applications of data science and machine learning across all sectors of the economy. Growth in artificial intelligence (AI) and cyber specialisms also drives demand for a broader supply of data skills, specifically at foundational level, as these help to feed the pipeline of advanced skills and provide businesses with the ability to work with data.

In 2015, Universities UK urged the higher education (HE) sector to 'do more' to embed data skills across all degrees.[5] However, according to research by Forrester in 2021, only half the UK's academic institutions have data literacy skills initiatives in place. With UK universities operating in a competitive international market, how data skills learning is embedded within or supplements curricula could increasingly inform the choices of overseas students and sponsors.

This all suggests that HE providers in the UK need to find ways to embed foundational data skills and data literacy across their programmes – not just in courses focused on data science – as almost all professions will utilise data in one form or another over coming years.

---

[1] https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy
[2] https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills-report.pdf
[3] https://www.gov.uk/government/publications/quantifying-the-uk-data-skills-gap/quantifying-the-uk-data-skills-gap-full-report
[4] https://www.tableau.com/sites/default/files/2021-06/Tableau_Data_Literacy_Report.pdf
[5] https://nanopdf.com/download/making-the-most-of-data-data-skills-training-in-english-universities_pdf

1.2 Programme aims and approaches

Through the Office for Students, the Department for Digital, Culture, Media & Sport (DCMS) provided funding in 2021-22 for an exploration of practice by universities in enabling HE students to develop foundational data skills.[6] The focus was deliberately on provision for undergraduates who are studying courses without a significant data science element ('non-cognate' students). The programme of exploratory work aimed to showcase and demonstrate some of the approaches being utilised in teaching these foundational data skills and, where possible, to gain any insight into which approaches work best. With those insights, other universities can potentially develop and embed approaches into their own provision that are seen as best practice, which ultimately should lead to more students at HE institutions in the UK learning these skills as part of a wide range of degree courses.

The Office for Students provided £294,000 of programme funding as part of an extension to the scope of the Postgraduate Conversion Course programme[7]. Funding was allocated as part of this programme to support seven UK universities to showcase their current provision of this type, while the Careers Research & Advisory Centre (CRAC) was commissioned to assist them in considering the effectiveness of their provision and identification of good practice. The seven 'pilot projects' received funding from October 2021 to March 2022 and concluded delivery of their funded activity in May 2022. This report presents emerging evidence from these pilot projects.

CRAC was also asked to undertake a wider evidence review to identify how other HE providers are teaching these data skills to non-cognate students (i.e. those on courses without a significant data science element) or supporting such learning, including outside the UK. This wider review took the form of an initial exploratory literature search, together with a web search exercise to identify provision of data skills modules and other learning by universities and/or third parties that was relevant to this project. The results of this wider review are also presented in this report. In practice, the initial literature search revealed very little evidence in the formal literature (as had been expected) so most of the evidence and practice that was identified was found through the wider, exploratory web-based search. However, a few published studies were found that contributed to the understanding of approaches being taken in certain countries. The wider evidence review was conducted in the early months of 2022, so this report is limited to evidence that was available in that respect up to 31 March 2022.

---

[6] By 'foundational' we do not mean learning specifically at foundation degree level (Level 5), but a range of data-related skills. These could include: management, modelling, cleansing and enrichment of data; data visualisation; statistical methods and data analysis; data quality assurance, validation and linkage; and some other skills within the broad concept of 'data literacy' (see section 2.1).

[7] Postgraduate conversion courses in data science and artificial intelligence – Office for Students

2 Evidence review

2.1 Scope: defining data skills and data literacy

The NDS makes explicit that there is as yet no consensus on what constitutes 'data skills':[8]

> *"There is no widely agreed definition of data skills. In this document we use the term broadly to cover the full range of basic, technical, governance and other skills – including project management, governance and problem solving – needed by practitioners to maximise the usefulness of data.*
>
> *The required technical skills range from programming, data visualisation, analysis and database management, to core skills such as problem solving, project management and communication."*

It also suggests that 'data literacy' is somewhat different, requiring "*some knowledge of data uses, some ability to assess the quality of data and its application, and the skills to conduct basic analysis*".

The evidence review revealed that such distinctions continue to be subjective. For example, the European DEDALUS study sees data literacy as a skillset that individuals increasingly will need in society and business to use data, and that is separate from more specific data skills:[9]

> *"Data literacy is the ability to read, write, critically assess, and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied — and the ability to describe the use case, application, resulting value, and its implications."*

An increasing number of countries are seeing data literacy and digital literacy as part of what they consider to be '21st century skills' that all individuals should gain through higher education, and studies are reflecting this understanding. A recent US literature review on data literacy education synthesised dozens of existing definitions and came up with its own definition as *"the ability to collect, manage, evaluate, and apply data, in a critical manner."*[10] Inspection of this in detail revealed that this did include data manipulation and 'computational thinking' skills that would be considered 'technical'.

In Singapore, the Ministry of Education has started a strategy through which all HE courses have a common element with the aim of "e*nhancing baseline digital competencies"* while selected courses aim to foster "*deeper competencies for certain sectors*".

---

[8] National Data Strategy - GOV.UK (www.gov.uk)
[9] *Data literacy: From higher education to the workplace*, DEDALUS, 2021. Home - DEDALUS European Project (cnr.it)
[10] *Strategies and Best Practices for Data Literacy Education*: Knowledge synthesis report, Dalhousie University, 2021

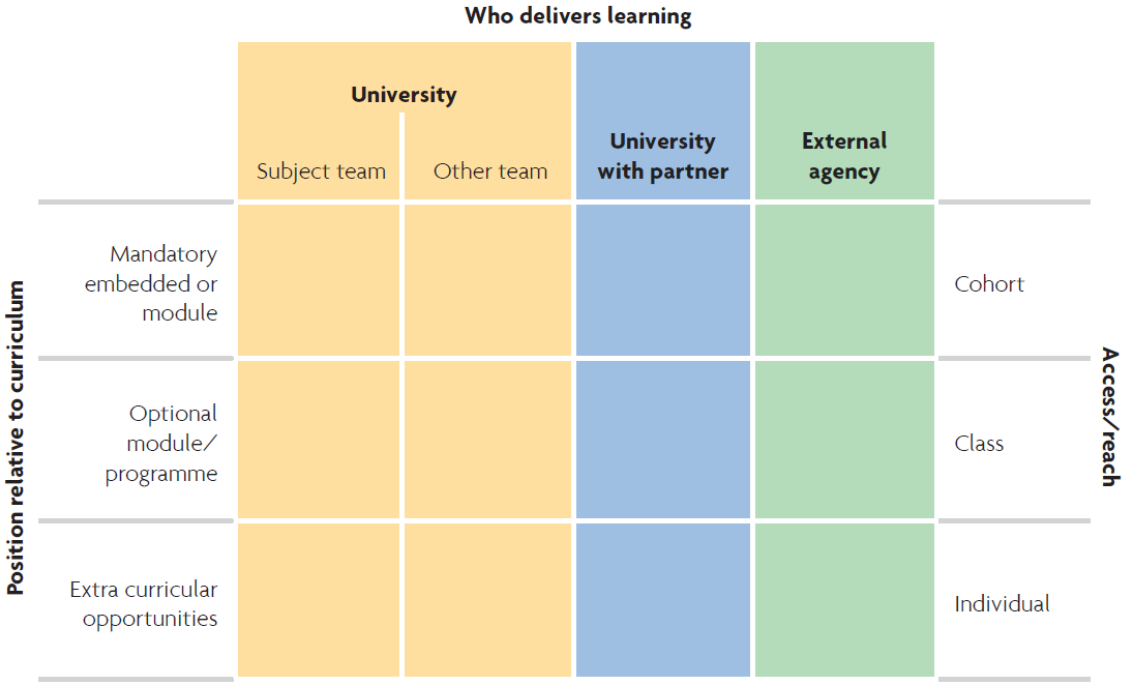
In the absence of full clarity of the distinction between what might be termed technical data skills and data literacy, we interpret the scope of this study – '(foundational) data skills' – to be a range of data-related skills which does include some of the more 'technical' data skills that are excluded in some definitions of data literacy.

2.2 A framework for approaches to developing foundational data skills

Initial analysis of examples of relevant provision found in the evidence review suggested it could be helpful to establish a framework with which different approaches could be illustrated and compared. Figure 1 outlines the framework developed as a product of this evidence review, which considers:

- how data skills provision is positioned in relation to the student's subject curriculum;
- who delivers that additional provision or learning;
- whether it is accessed by the student on a cohort (class) or individual basis.

Figure 1: Framework for approaches to development of foundational data skills



In relation to the first of these dimensions, the position of provision relative to the curriculum, it quickly became clear that in the examples of provision identified in the evidence review, some of the teaching of data skills was in the form of a discrete module or other unit of taught content that was a mandatory part of the course curriculum. In turn, this could be delivered by faculty-based staff teaching the curriculum, in which case the examples used were typically set within the subject context, or could be delivered by data science or computing specialists from another department (commonly referred to as a 'service teaching'

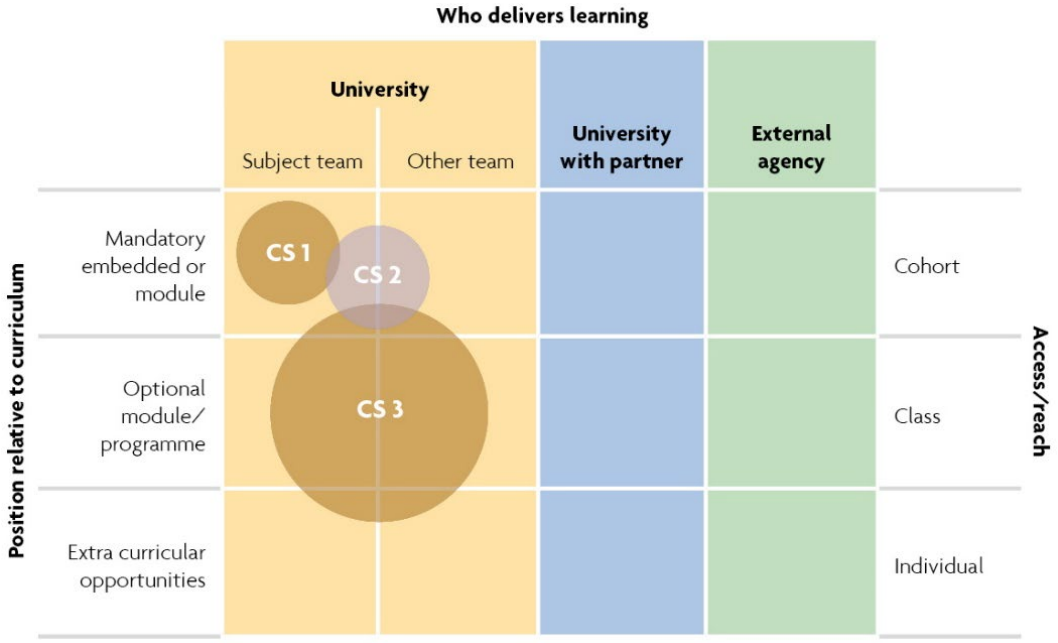
model). In many cases, this latter model meant that examples were not customised to the context of the degree subject. Alternatively, data skills learning could be an optional or elective module, not a mandatory part of the curriculum, and again taught either by a subject or service teaching model. There were also approaches where data skills learning was outside the curriculum entirely, in many cases utilising third party content and providers.

These examples immediately demonstrate that while there is a separate dimension of 'who delivers' the teaching, it is not entirely independent from the location of the teaching in relation to the curriculum (i.e. whether fully embedded or not). We also realised quickly that there was a related dimension in terms of the mode through which students access data skills learning (or, viewed in a different way, how far the provider's learning reaches out). Thus, modules within the curriculum are generally accessed within a cohort setting (i.e. learned by the student in a class), whereas certain other types of provision are accessed by the student individually, such as through the university's online learning platform. Access to such provision from third party providers can also be offered to individuals much more widely, independent of the university, as students and university staff are simply two potential customer groups.

2.3 Examples of approaches

In this section we present a range of case studies of different approaches to development of data skills amongst non-cognate HE students. These are arranged according to segments of the framework in Figure 1. In addition to one or two case studies illustrating each type of approach, some further specific examples are listed which may be worthy of further investigation by readers in particularly relevant contexts. Given the extent and range of practice, these lists are illustrative for a range of non-cognate disciplines and contexts, rather than exhaustive.

2.3.1 Mandatory embedded module/s



In this first section, we highlight a case study from the University of York, which focuses on a mandatory embedded data skills module, and two case studies of broader approaches from outside the UK.

Case Study 1: University of York
Big Data Biology[11]

This core online module takes place in the second semester of year 2 and is worth 10 credits. It is described as being designed to "*extend student capability with [the programming language] R and statistics and [will] teach biological concepts, analysis and data handling skills that will be relevant to a variety of other courses*". Its learning outcomes are:

1. Explain the sources, value and caveats of large data sets in biology.
2. Apply appropriate data analysis and visualisation to large biological data sets.
3. Be able to organise large data sets and document computational analysis.
4. Interpret large data sets using online resources.

On the module's welcome page, coordinator Dr Daniel Jeffares reminds the student of its core purpose: "*This module will teach you how to analyse large biological data sets. We use RStudio as a tool, but this is not a module about RStudio. We use statistical tests, but this is not a module about statistics.* ***This module is about using large tables of data to understand biology.*** *Keep your eyes open, you may see something wonderful.*"

Students are also reminded that there are potential employability benefits: "*Your RStudio data analysis skills can be used in all kinds of jobs. So mention your proficiency in data analysis on your CV. RStudio is free, so you can always have a copy*."

The module is structured as follows:

- The module begins with two recorded video lectures introducing data analysis and RStudio.
- In weeks 2 and 3, students watch videos about: (1) the course datasets; (2) data analysis concepts; and (3) RStudio skills. They then practise R skills in RStudio.
- An online workshop programme is introduced in week 4 with videos followed by exercise practice. For each 90-minute exercise, students choose two data sets that they find most interesting.
- In weeks 5-8 students watch videos and attempt exercises before attending a weekly Zoom workshop to work through the exercises with staff.
- For assessment, students must do some analysis using one of the data sets provided and write a 1500 word report about it. To support this, staff offer Zoom drop-in sessions and Q&A sessions in weeks 9 and 10. Assignments are submitted in week 1 of the summer term.
- An optional fifth Zoom workshop is offered in summer term week 8.

The presentation of the module and resources aims to support student confidence and engagement using a direct, approachable style and making orientation straightforward. Ongoing support for difficulties encountered also includes a Google hangout and virtual learning environment (VLE) discussion board.

[11] Big data biology (BIO00047I) 2022-23 - Module catalogue, Student home, University of York

Additional examples:

- University of Bath, Economics: (Core Skills for Economics: Statistics & Data Analysis)[12]
- University of Glasgow, Psychology: (1a – social, developmental, health & wellbeing)[13]
- University of Leeds, Geography: (Digital Geographies)[14]
- University of Southampton, Marine Biology: (Data Literacy and Analysis)[15]
- University of Sussex Business School, Business & Management Studies: (Digital, Data and Decision-Making Skills)[16]
- University of Sussex, Psychology: (Research methods - four modules across years 1 and 2)[17]

The UK's Q-Step programme, funded by the Nuffield Foundation and Economic and Social Research Council, could be considered to have had broadly similar aims to an increase in foundational data skills. Q-Step was developed as a strategic response to the shortage of quantitatively skilled social science graduates in the UK and designed to promote a step-change in quantitative social science training through the development of exemplar provision. Broadly, it is a funded programme with a network of Q-Step Centres that deliver courses with enhanced quantitative skills learning (developed through Q-Step funding). A recent evaluation concluded that it had broadly been successful, that Q-Step centres should continue to learn by sharing practice, and that universities and funders should consider whether the model could support development of quantitative skills in other subject domains.[18]

Outside the UK, the development of foundational data skills in universities and polytechnics in Singapore is reported to have advanced at pace as a result of an integrated planning system including refreshing its HE curricula. This was based on an ambition both to enhance baseline digital competencies amongst all students (including digital wellbeing and ethics) and to deepen those and related competencies for those studying subjects relating to sectors ripe for AI adoption (such as cyber-security, logistics, manufacturing and finance).

Nanyang Technological University (NTU Singapore) introduced a revised common curriculum for all first-year students with compulsory courses developing key transferable skills including 'Navigating the digital world'. That course focuses on how to apply computational thinking and reasoning, solve problems and analyse data, as well as teaching online ethics and helping students identify online threats and fake news. Students take the course in multidisciplinary groups. Deeper learning in data skills is discipline-specific, and

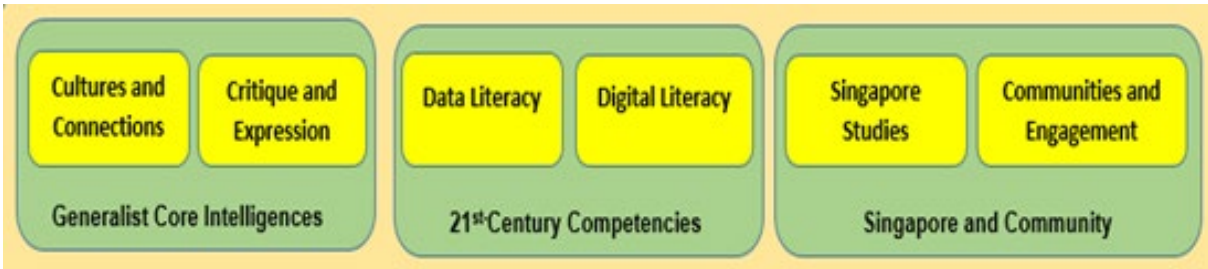
[12] Programme & Unit Catalogues - University of Bath
[13] Microsoft Word - 2020-21 L1 PSYCHOLOGY HANDBOOK.docx (gla.ac.uk)
[14] Module and Programme Catalogue (leeds.ac.uk)
[15] Data Literacy and Analysis | SOES1015 | University of Southampton
[16] Digital, Data and Decision-Making Skills (N1629) : Business and Management Studies : ... : University of Sussex Business School internal site : University of Sussex
[17] Research methods : Teaching and student experience : School of Psychology : University of Sussex
[18] *Evaluation of the Q-Step programme*, Technopolis Group report for the Nuffield Foundation, 2022

can involve external partnerships, such as with Tableau Software. Examples of how this works in practice can be seen in course descriptions.[19]

Case Study 2: National University of Singapore

In 2015, National University of Singapore (NUS) made a 'quantitative reasoning' course mandatory for all first year students, as part of a general education curriculum (which has subsequently been revised). That curriculum now comprises six pillars, of which data literacy and digital literacy are two (see Figure 2). Students are required to take one module from each pillar.

Figure 2: Enhanced general education curriculum NUS[20]



Within this structure, a Computational Reasoning module teaches coding and data analytics, together with more philosophical critical thinking skills to assess the ethical, social and political implications of these computational methods. A compulsory Data Literacy module is taught to 800 arts and social science undergraduates each semester.

Learning is through a combination of online lectures and classroom sessions: "*Besides learning to analyse, interpret and use data to produce reliable results, students… [were]… exposed to how data is gathered and used in various contexts, such as in health and population*." NUS states that students often come into the module with psychological barriers to effective learning, as a result of anxiety facing an unfamiliar subject and fear that the module could affect their overall grades. Interactivity, such as quizzes and group projects focused on real-world challenges, as well as humour, are used to alleviate those anxieties. One example is students learning to code by developing a maze game in Excel, which teaches the same mechanics as other algorithms but in a more relatable way.[21]

[19] E.g. https://web.nbs.ntu.edu.sg/undergrad/common/contents/courseoutline/co_AC2401.pdf

[20] https://www.nus.edu.sg/registrar/academic-information-policies/undergraduate-students/general-education/for-students-admitted-from-ay2021-22

[21] Teaching data analytics to arts students

Case study 3: University of British Columbia, Canada
EDUCE

The Experiential Data science for Undergraduate Cross-Disciplinary Education (EDUCE)[22] initiative in the Faculty of Science of the University of British Columbia (UBC) aims to build data science competency by integrating data science modules into both mandatory and elective courses, supported by additional co-curricular activities. Its learning objectives require that students learn to: recognise and define uses of data; explore and manipulate data; visualise data in tables and figures; and apply and interpret statistical data.

EDUCE modules are designed to be delivered as standalone classes, or as separate recurring items (e.g. 'data science Fridays') or to be integrated in consecutive fashion, all hosted on the university's GitHub. Each module uses discipline-specific data sets, questions and software. Early modules focus on introduction and practice with pre-cleaned, tidy data; later modules enable application of data science skills to higher dimensional and un-cleaned data, focusing on visualisation, interpretation and communication. The modules are supported by learning opportunities outside the classroom – workshops, hackathons and directed studies. The teaching team spans 10 departments across three faculties.

Example: Undergraduate microbiology

EDUCE modules have been integrated into seven third and fourth year microbiology courses, although in future the intention is for modules to be positioned earlier. Students on different courses have access to modules based on the programme requirements, while the co-curricular activities to some extent compensate for those differences.

In 2021, 475 students were taking part in EDUCE modules annually, while undergraduates made up around 20% of all participants in co-curricular workshops. Preliminary analysis of student feedback suggested that participation impacted positively on students' interest in topics such as bioinformatics.

In Germany, regional approaches have been taken rather than national, although in recent years a nationwide Data Literacy Education Network[23] has been funded, comprising over 20 universities. Through working and topic groups, experiences are shared, there is exploration of what works best, and joint development of solutions including open access resources. Examples of German universities' approaches include:

- Georg-August University of Goettingen developing a Data Literacy Basics course for all undergraduates, establishing a 'DataLab' as an interface between various disciplines and organisations supporting the regional economy, and curating a collection of open educational resources that can be credit bearing with a form of examination.
- At the Leuphana University of Lueneburg, data literacy education is embedded in its study model. The 'Leuphana Semester', with which all students begin, teaches mostly interdisciplinary content but with project work to develop basic skills in mathematics,

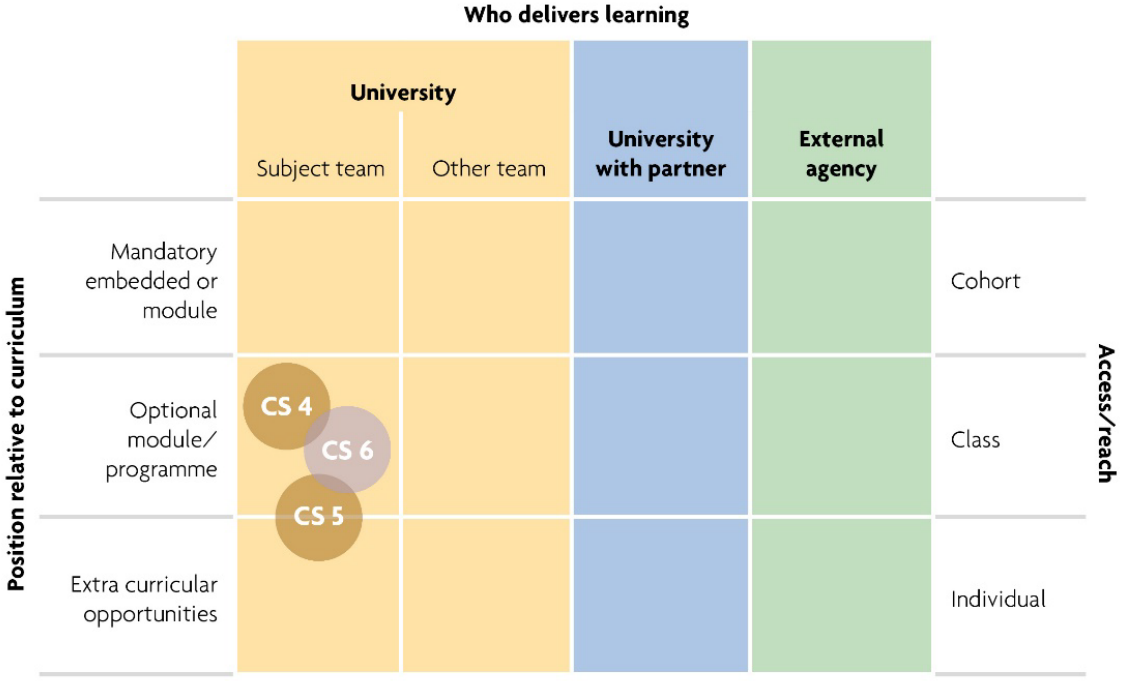
[22] An integrated, modular approach to data science education in microbiology. - Abstract - Europe PMC
[23] Data Literacy Education | Stifterverband

statistics and programming. Students later move onto more independent practical data analysis and visualisation projects in groups, using real data from the university's regional partners.

- A joint project between the University of Bielefeld, University of Paderborn and University of Applied Sciences Bielefeld aims to develop a data skills certificate tailored to the regional economy and science, based on a data literacy competency framework, to accredit its data-literate graduates.

2.3.2 Optional modules within a programme

In this section, we highlight a number of case studies and examples where data skills are taught as optional modules within a non-cognate study programme.



Case Study 4: University of Manchester
Learning data skills in the classroom and workplace (social sciences)

Through the Manchester Q-Step Centre[24] undergraduates "*learn techniques to help understand, analyse and criticise data through an integrated set of new courses and work placement opportunities*". This case study illustrates two, linked approaches to data skills learning at the University of Manchester:[25]

1. A second-year course unit, 'Making sense of criminological data', developed as part of the Q-Step Centre programme
2. 'Data Fellowships', a work placement initiative.

[24] About Q-Step - Faculty of Humanities - The University of Manchester
[25] Carter, J. (2021). Developing a future pipeline of applied social researchers through experiential learning: The case of a data fellows programme. *Statistical Journal of the IAOS*, 37(3), pp. 935-950

Completion of the course unit is mandatory for any criminology students who wish to apply for a Data Fellowship but is otherwise optional.

'Making sense of criminological data' is the first of two second-year units developed by the Q-Step centre and takes place in the first semester (the other, 'Modelling criminological data', is taught in the second, using R). It builds on the first-year mandatory unit 'Criminological research methods' which introduces quantitative and qualitative methodologies and research design.

The unit provides an opportunity for second year undergraduates to develop their data analysis skills through examining topics of interest to the lecturer. All the practical materials have been made available online through GitHub. The delivery mode is a combination of lab classes, where students work in groups with the support of the lecturer and a Graduate Teaching Assistant (using remote teaching in 2021), and lectures. The unit uses Excel for the quantitative analysis, and students analyse data from the Crime Survey for England and Wales, and some local police force data.

Manchester's Data Fellowship programme was designed at the outset of the Q-Step initiative in 2013 and by 2021 had created some 300 paid work placement projects in over 60 organisations in industry, government and the third sector, for students to practise their data and statistical skills in the workplace. Eligible students apply to advertised opportunities lasting up to eight weeks in the summer vacation following their second year.

Students on these placements are encouraged to use tools developed for the programme: frameworks for reflection and two sets of tools to help students monitor their progress, including with developing data skills.[26] Recognition comes in the form of an annual conference featuring poster and presentation competitions. Much of the value of the work placement is in preparing students for careers they may enter after graduating.

Jackie Carter, Professor of Statistical Literacy, stated that the Data Fellowship programme has "*informed the development of the undergraduate curriculum and enabled reflection on the skills and software that we teach. Data Fellows are graduating into careers in fields that would previously have been difficult to enter without a STEM (Science, Technology, Engineering and Mathematics) degree. To date, 70% of Data Fellows are female, with 25% from disadvantaged or under-represented groups*."

As this example from the University of Manchester shows, optional data modules that support an academic programme may be combined with workplace-relevant learning. (This is also true of the University of Exeter example from Digital Humanities mentioned below.) An alternative model that focuses primarily on employability support is outlined in the following case from the University of Bradford.

[26] Carter, J. *Work Placements, Internships & Applied Social Research* (2021). London, Sage Publications. Chapter 7

Case Study 5: University of Bradford School of Management
Data skills learning with industry partners

The School of Management was named Business School of the Year at The Times Higher Education Awards 2021 and the Educate North Awards 2020. The school's submission to the latter included its 'Career Booster' programme, which enables students to learn industry-relevant skills and take part in vocational training in areas such as AI and Big Data, psychometric testing and mock assessments.

The Career Booster programme spans the Faculty of Management, Law and Social Science. Students may choose to timetable up to two weeks each academic year on the programme. In 2020 the programme provided 7,500 hours of activities overall, representing an increase of 35% on bookings compared with the previous year.

Within Career Booster options on 'Big Data analysis and reporting', students may choose analytical workshops such as:

- Amazon Web Services Cloud workshops
- Introduction to Python Analytics
- SAS Enterprise Querying and Reporting

Through these workshops, students gain hands-on experience of accessing, managing, summarising and analysing data from different sources, and presenting results in tables and graphs. Students booked onto the SAS workshops are also encouraged to access free SAS cloud-based software,[27] which is a browser-based programming environment using SAS Studio, where they can learn SAS programming from basic to advanced techniques using coding or point-and-click tasks.

Development of a curriculum that attempts better to meet workforce needs is the focus of the final case study in this section, from the University of Gloucestershire.

Case Study 6: University of Gloucestershire
Data skills for international development careers

The university was the lead partner of a European Union Erasmus+ project 'Development Counts: Data Skills for International Development Careers'. Among other activities, a 2021 pilot project has supported students in social science disciplines at four partner universities to develop their data skills by completing 'real life' international development data projects in cross-country teams.

The five-day pilot training programme involved 16 students from the University of Gloucestershire, Gazi Üniversitesi, L-Università ta' Malta and Universidad Nacional de Educación. Activities varied from mastering new software to taking part in online data scavenges, and were designed to develop both data skills and skills in teamwork, intercultural awareness and collaboration. In addition, students worked in international teams to conduct and present projects on development data blind spots.

[27] SAS OnDemand for Academics | SAS UK

The design of the programme was informed by interviews with global development professionals in Malta, Spain, Turkey and the UK. These highlighted that technical data skills alone are insufficient, and that successful use of data for good in global development requires graduates to be able to critically evaluate data and its quality, navigate the politics of data, and collaborate with and influence a range of stakeholders.

The programme sought to provide foundational training for non-specialist students and to build confidence and enthusiasm around working with data, focusing on five key elements:

- The current and future use of data skills in global development;
- Sourcing data;
- Visualising data;
- Communicating data and using data for decision making;
- Ways to further develop data skills and development knowledge.

There were three parts to the training programme:

1. An online self-study course in five modules to introduce key material.
2. Online live tutor-led activities to extend and consolidate learning from self-study modules and support team building.
3. Activities to support students complete remote team projects, applying their learning to real-world scenarios whilst developing cross-cultural teamwork and communication skills.

The training programme was designed for groups of students from different institutions and countries to complete together online over a one-week period. However, the learning resources can also be mixed and matched for other appropriate groups of students. The resources are freely available from the 'Development Counts: Data Skills for International Development Careers' website.[28] There is also a function where universities interested in running the programme can register to be matched with other interested universities to jointly run the programme.

Additional examples of teaching data skills through optional modules include:

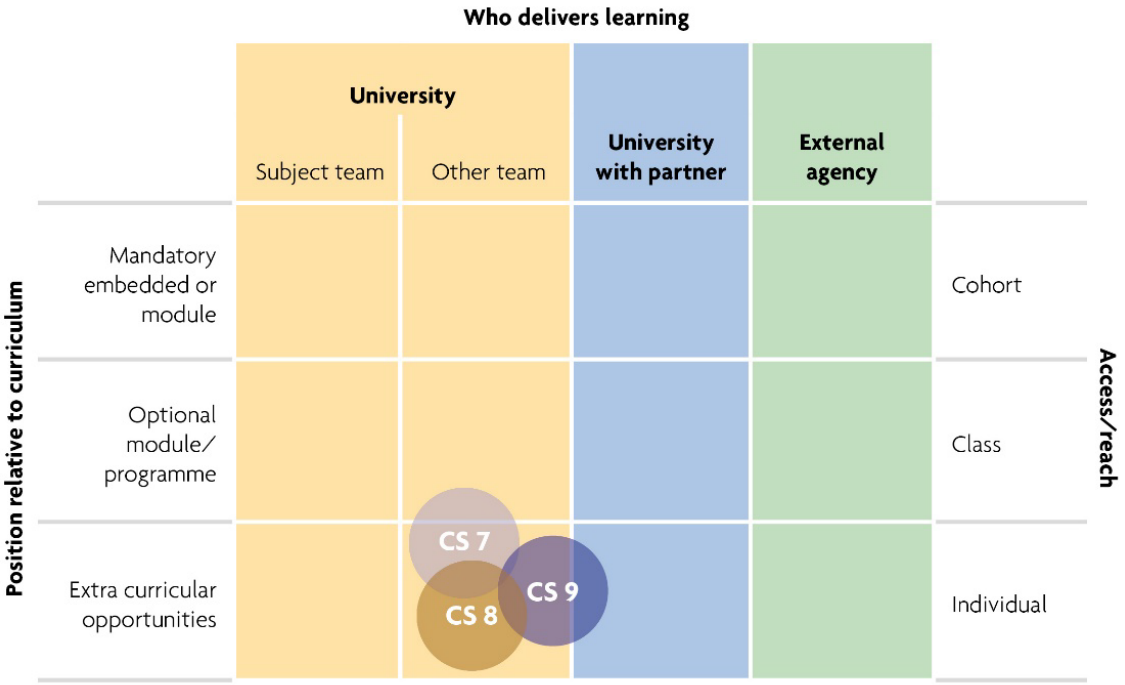
- University of Bradford School of Management, e.g. BSc Accounting & Finance and BSc Business & Management, Year one (Fundamentals of Artificial Intelligence & Data Analytics);[29]
- University of Exeter, Liberal Arts/Digital Humanities: Data Analysis modules,[30] including an assessed work placement. Note that although 'optional', the choice of options is limited.
- University of Warwick, Liberal Arts: Quantitative Methods for Undergraduate Research module [31] (an 'innovative' problem-based learning approach).

[28] Training Programme | Development Counts: Data Skills for International Development Careers (uniofglos.blog)
[29] Accounting and Finance BSc (Hons) degree - University of Bradford
[30] Modules | Liberal Arts | University of Exeter
[31] IP110 Quantitative Methods for Undergraduate Research (warwick.ac.uk)

Universities are increasingly partnering with external organisations to supplement and enhance the in-person teaching they provide in data skills. Examples of such third party provision are included in section 2.3.4.

2.3.3 University extra-curricular modules/support

Here, we highlight several case studies, in and outside the UK, which provide examples of extra-curricular data skills modules or programmes of support.



Case Study 7: University of Edinburgh
Data Skills Programme[32]

University of Edinburgh Information Services offers a six-month blended programme – 'Developing your data skills' – for staff and students to complete at beginner, intermediate or advanced level.

It is framed around the Jisc Digital Capability Model[33] (data skills included in the Information, Data and Media Literacies competence) and based upon freely-available online self-study resources verified and collated into a step-through programme on the University's VLE. Each level is designed to take around 1-2 hours per week over a six-month period to complete.

The online programme is supplemented by four optional workshops to help embed learning. Included is the opportunity for participants to undertake a project in which they access a dataset, cleanse and analyse it, and show the results in a group presentation. Participants who undertake the optional project gain a certificate of course completion.

[32] https://www.ed.ac.uk/information-services/help-consultancy/is-skills/programmes-courses-and-toolkits/development-programmes/data-skills
[33] JFL0066F_DIGIGAP_MOD_IND_FRAME.PDF (jisc.ac.uk)

For each level ('pathway'), content is divided into: Introduction to Data Skills and Programming; Statistics and Data Analysis; and Presenting Your Data. At the beginner's level this includes:

- Basic understanding of data science.
- Basic programming in Excel and R, Python or SQL.
- Introduction to statistics (up to around a beginner's statistics module on an undergraduate course).
- Introduction to data analysis and visualisation skills.

Undergraduate uptake appears to form only a small proportion of total participation in the programme. When we enquired, the 2021/22 programme was expected to have only one of twelve cohorts formed dominantly of undergraduate and postgraduate students, so our understanding is that the majority of participants are professional services staff and researchers.

In contrast to the self-standing data skills programme offered by the University of Edinburgh, above, provision by University College London (our next case study) is based on a menu of discrete sessions.

Case Study 8: University College, London (UCL)
Digital Skills Development

UCL's Digital Skills Development (DSD) team develops the digital skills of UCL's staff and students to meet their academic, professional and personal needs, ranging from support for desktop and office functions through to data analysis. While pre-pandemic most of this provision was face-to-face, most has shifted online and now only a small proportion is campus based. Individual support is also offered through online learning and (currently virtual) drop-ins. The team's web pages list courses by theme,[34] including Data Analysis, Mastering Excel and Programming (which have some overlap). One-hour demonstration sessions are also offered for students embarking on dissertations or theses. These are pitched as: "*geared not towards teaching you to use software apps, but to understand a task associated with learning or research, to know what software is available and how to choose the right app for your work"*.

Options include a session 'Software for success: data analysis and statistical tools', which covers the following:

- What do you need from a data analysis package?
- What tools does UCL make available and support?
- Which is the best tool for you?
- Mini demonstrations of SPSS, Stata, MATLAB and Python.
- Code development.

The short courses offered by DSD live online (using Blackboard and Microsoft Teams) recently included the following:

[34] Digital Skills Development courses by theme | Information Services Division - UCL – University College London

- An introduction to R with RStudio: 9 hours across four remote classroom sessions.
- Getting started with Stata: 8 hours across four remote classroom sessions.
- Data visualization in R with ggplot2 (half day).
- Data manipulation in R with RStudio (half day).

As noted in the Edinburgh case study, the extent of uptake of this provision by undergraduates is low compared with postgraduates and university staff. We understand that undergraduate students comprise under 3% of total uptake, compared with postgraduate uptake of nearly half.

Case Study 9: University of California, Berkeley
Data 8: The Foundations of Data Science[35]

Data 8 is a course open to all undergraduates at Berkeley, offered by its Division of Data Sciences. Course enrolment is reported to have grown from under 100 in the 2015 pilot to over 1,600 by early 2019.

Data 8 combines three perspectives: inferential thinking, computational thinking and real-world relevance. Students use real data to understand relationships and patterns while being taught critical concepts and skills in computer programming and statistical inference. A 15-week course, Data 8 consists of in-person lecture sessions (also available online) and labs with homework assignments as well as mid-term and final examinations. The curriculum and format were designed specifically for students who have not previously taken statistics or computer science courses; 'high-school algebra' is the only stated prerequisite.

In 2019 the University summarised Data 8's achievements, highlighting: *"its success equipping a broad range of students, including those without math or computing backgrounds, with a strong foundation in just 15 weeks. The course accomplishes this by initially focusing on concepts like inference and random sampling with real world data rather than first delving into math and computational theory*".

Data 8 uses the Jupyter cloud computing infrastructure, which allows students to access computing tools without loading and launching programs, and it takes examples from a variety of disciplines, from biology to sports and literature. Additional 'Connector' courses[36] that accompany Data 8 enable students to explore data science further in their area of interest.

Students from under-represented groups have access to an additional summer course for incoming students and a linked Data Scholars program providing mentoring and research opportunities. Roughly half of the students who enrol on Data 8 are reported to be women.

All course materials, including textbook and assignments, are available free online, and there is an online course variant (Data 8X) for self-paced study through EdX.

A number of other universities have developed customised local versions of Data 8 including Yale, Cornell, University of Chicago, NYU, UC San Diego and UC Santa Barbara.

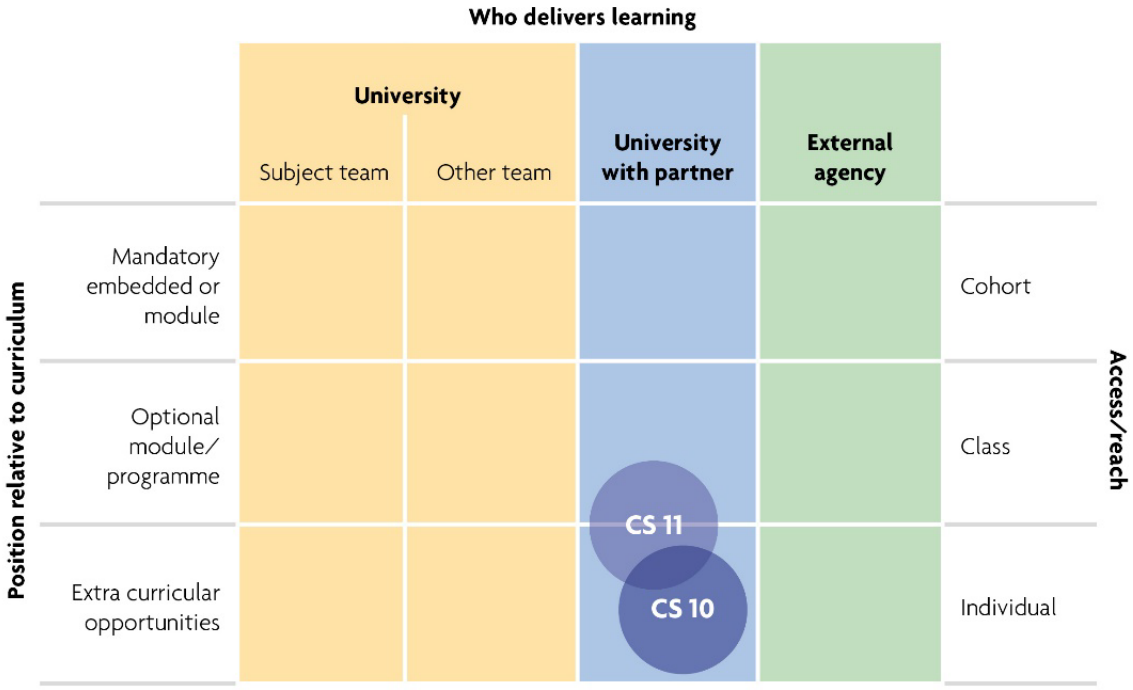
[35] http://data8.org/

[36] Data Science Connector Courses | CDSS at UC Berkeley

Other examples of campus-wide data skills programmes in the UK include:

- University of Exeter, Q-Step Centre broader provision, e.g. Pathways to Data Analytics.[37]
- University of Manchester, University College for Interdisciplinary Learning: Trust & Security in a Digital World.[38]

2.3.4 Provision supported by a third party



Based on the framework in Figure 1, we found examples of external agencies (third parties) involved either in supporting the provision of learning by a university or directly providing learning opportunities to individual students. In this section the former of these strategies is considered. In many cases it will be in the form of generic external support rather than being tailored to the student's degree course programme.

One example already covered is where a university mediates a course provided by another university, such as other US universities' adoption of Berkeley's Data 8 course (Case Study 9), which may or may not include some customisation by the local university.

In the UK context, there are a number of examples where universities are utilising and promoting/facilitating access by their students to external programmes, both online and on campus.

- The Nuffield Foundation-funded Q-Step programme is currently helping 17 UK universities to run Q-Step Centres that support development and delivery of undergraduate programmes in social science. These have the specific aim of delivering

[37] Data Analytics | Career Zone | University of Exeter
[38] UCIL20132 - Manchester - Trust and Security in a Digital World: From Fake News - StuDocu

quantitative skills training to give social science students a deeper and more secure grasp of quantitative methods for the evaluation and analysis of data. In addition to the degree programmes run under the Q-Step banner, its centres also provide a range of summer schools as well as facilitating summer research placements and internships. There is some overlap between the content of these programmes and the data skills learning considered in this report.

- Amongst many other universities, University College London facilitates access for its students and staff to LinkedIn Learning, through its Information Services Division. [39] LinkedIn Learning offers a vast array of online self-study courses (c.6000). Within its 'Technology library', there is a section on data science, which includes courses on themes such as big data, data analytics, data modelling, data governance, data visualisation, as well as machine learning and AI. In addition to facilitating access to these courses (which would incur fees if accessed individually), UCL sets up learning groups through which students or staff can share tailored content and keep track of their progress.

Case Study 10: Birmingham City University
Qlik Academic Program [40]

The Graduate+ programme is an extra and co-curricular award framework offered to students at Birmingham City University (BCU), with the aim of enhancing progression and employability. Within this framework, BCU is offering the Qlik Academic Program free to students on any BCU course.

Qlik gives students access to a range of software and training courses including role-based courses such as Business Analyst and Data Architect. There is a data analytics curriculum which provides lecture notes and course materials, as well as videos and student exercises. These aim to provide a grounding to concepts and theories within data analytics, and potentially data literacy certification. These are all offered through Qlik's 'continuous classroom' online learning platform. Qlik claims that it has students in more than 1600 universities in 85 countries worldwide.

Behind its online learning programmes, Qlik is a commercial software suite and platform that aims to help businesses, organisations and customers to integrate and analyse business data. It is one of a number of software providers that offer learning programmes to those using its products.

The Bright Initiative is similarly an offering from a data platform and software provider. Its academic programme aims to offer its technology for use in data-driven social research by those in HE and others, backed by a range of workshops and programmes which develop data skills and expertise using its platform and tools. It claims to partner with 150 academic institutions worldwide, including some leading UK universities.

[39] LinkedIn Learning | Information Services Division - UCL – University College London
[40] qlik-education-academic-program-brochure-en.pdf

Case Study 11: The Carpentries

The Carpentries[41] is a project by a non-profit organisation in California, bringing together three organisations that historically have supported the teaching of computing and data skills to researchers. The Carpentries aims to be an inclusive community of individuals who teach or need to learn computing and data skills, with the broad objective of capacity building amongst researchers. A major strand of its activity is development and provision or facilitation of domain-specific workshops on the fundamental data skills needed to conduct research. The target audience is researchers with little or no prior computational experience; learners are expected to be graduate students or others at postgraduate level. The current range of workshops includes:

- Astronomy
- Ecology
- Genomics
- Geospatial data
- Social sciences

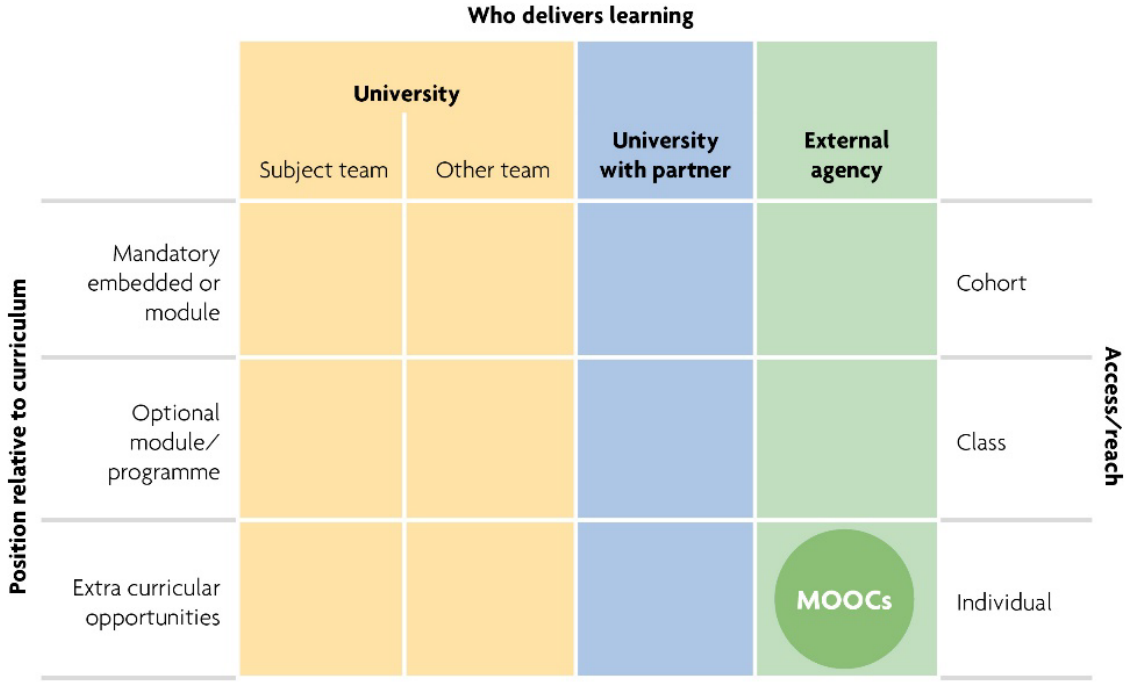
Further domains under development include image processing and economics. Although not a workshop in the same way, there is also a biology curriculum available for use by those teaching data skills in this domain.

For each workshop it has developed a curriculum and series of lessons. In addition to the curriculum, it provides datasets, tools and instructor notes, and the content is maintained and kept up to date by a range of specialist individuals ('maintainers'). It is claimed that the materials assume no prior knowledge about the tools, and in principle the workshops are designed for non-cognate learners.

Many of its workshops are online and in the majority of cases are hosted by a partner university or education organisation, whose students or researchers are eligible to participate. In some cases there is more open eligibility (for example, the UK Science and Technology Facilities Council is hosting a Software Carpentry workshop in autumn 2022 which will be open to relevant graduate students and researchers).

[41] The Carpentries

2.3.5 ‘External’ (third party) provision



In this last category, we consider external opportunities which a university may promote to its students but which do not have any other involvement by the university. To some extent the examples in this category and the previous one (section 2.3.4) overlap, as an external provider may offer both university-mediated and unmediated access. The potential range of activity within this category of provision is very wide and also international, ranging from specific courses which lead to qualifications or certification, typically provided through MOOC[42] platforms or similar, to collections of resources or other information services.

It is also worth noting the offerings of bodies and organisations which provide courses and resources primarily for people already in the workforce, rather than aimed specifically at students, because students too could presumably access such resources. We should also remember that some students study part-time and could be employed at the same time, which could mean some of those ‘professional’ resources are appropriate for them. The UK’s Office for National Statistics Data Science Campus is an example of a ‘professional’ offer of support, including mentoring and individual continuing professional development (CPD) modules at Masters’ level.[43]

A number of MOOC platforms provide online self-study courses in data skills and on related topics, including Coursera,[44] OpenLearn,[45] Future Learn[46] and edX.[47] For example, edX offers well over 100 courses under its ‘data analysis’ heading, which are from a wide range

[42] Massive Open Online Course - a free online distance learning programme which aims to facilitate large-scale participation to individuals through flexible and open access to learning materials.
[43] Learning and development | Data Science Campus (ons.gov.uk)
[44] Coursera | Degrees, Certificates, & Free Online Courses
[45] Open Learning - OpenLearn - Open University
[46] Online IT & Computer Science Courses - FutureLearn
[47] Learn Data Analysis with Online Courses, Classes, & Lessons | edX

of providers including US universities, such as Harvard, Berkeley and MIT, and the University of Edinburgh, as well as companies like IBM. The model tends to be free access to course materials but a fee if an examination and certification are sought.

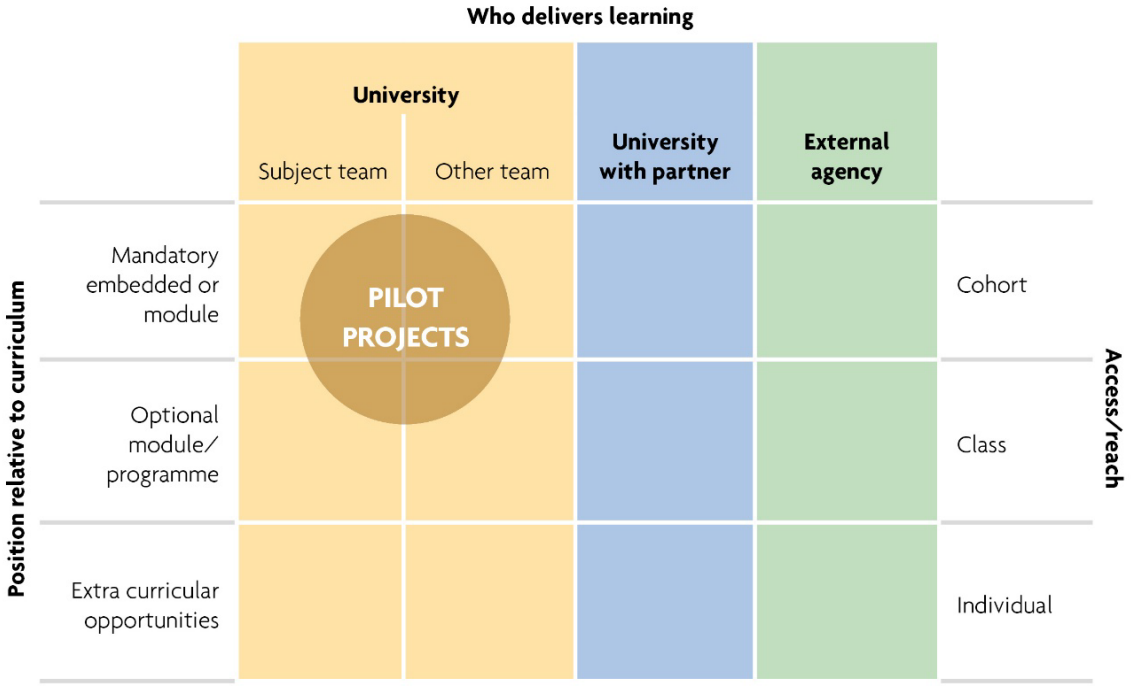
2.4 Summary

This evidence review has highlighted a number of examples of pedagogical practice in the development of foundational data skills, for non-cognate as well as cognate students, of which many are at undergraduate level. We have developed case studies from the UK and beyond which showcase a range of possible approaches to this, and have categorised them within a wider framework. What is largely absent, however, from the evidence is robust insight into the effectiveness of these approaches. There is some evidence about uptake, and this suggests some of the wider approaches have been not embraced by large numbers of undergraduate students.

In the next section of the report, we provide details of the seven UK HE providers which received funding to undertake pilot activities in relation to approaches to teaching data skills at their institution to non-cognate students – defined here as those on courses without a significant data science element. The approaches considered in the seven pilot projects can also be mapped within the framework. The descriptions that follow in the next section provide some more detailed insights into how these universities are beginning to evaluate their current strategies for foundational data skills provision.

3 Pilot projects

The seven pilot projects funded through the programme almost all focused on mandatory modules, delivered by either the subject team for the course or using a service teaching model – hence they are plotted in the upper left segments of the framework.



3.1 Solent University

Solent's project was predicated on its understanding that many students enrolled on non-cognate courses do not expect to have to learn specific data skills, as they believe them to be taught only within cognate courses, and that there is greater disparity in levels of computer literacy among non-cognate students than cognate. The project was located in its BSc Business Management and BSc Accounting and Finance courses, which both have a Level 4 'DATA' module (Data Analysis, Tools and Applications) in semester 2. Currently, the first block of the module is taught using collaborative and enquiry-based learning, while the second block uses 'flipped classroom' and reflective learning.

There is a student profiling exercise right at the start of induction to the courses (which assesses levels of digital skills) and early in the courses an optional self-paced module on foundational data skills, which runs through semester 1. Activities within the latter were designed specifically for non-cognate course students and use data and examples from their subject domain to demonstrate the benefits and applications of data skills. Students earn badges for completing these mostly video-based activities, while tutor-assisted discussion forums and weekly online support sessions assist students but also monitor and evaluate their progress. Two formative assessments are undertaken during semester 1.

The topics covered in Block 1 of the DATA module include: Introduction to data analysis; foundations of data analysis; data collection; data preparation; descriptive statistics; and

data visualisation. During lectures (live or online), students participate using Mentimeter for quiz questions or polls via their own devices. In practical sessions on-campus, groups of learners work together to solve problems.

In Block 2, topics include: probability concepts; cluster analysis; regression analysis; time series analysis; and forecasting and data for decision-making. Students are provided with short, pre-recorded lecture segments (with embedded quizzes) to watch before each class. Digital badges are awarded to incentivise participation. Independent reading is necessary to understand the topic in detail. Students also describe their learning experiences through short self-reported questionnaires for the reflective aspect of the pedagogy. This approach helps students to be aware of their own learning process and analytical capabilities.

Project activities

Student surveys were deployed before and after the DATA module and feedback was also obtained from surveys on each block. These gathered data about attitudes, confidence levels and self-assessment of their quantitative and digital literacy skills. Results from the formative assessments in semester 1 and summative assessments in semester 2 were also used for the project, set in the context of the pre-course student profiling.

In total, 140 students completed the DATA module, including 117 from the two non-cognate courses. 80 students had taken part in the profiling exercise, including 71 responding to questions on data literacy and 67 on digital skills. However, only 19 students responded to both the semester 2 pre- and post-DATA module surveys.

Findings

Results of numerical testing before and after the module were compared for cognate (computing) students and also the target non-cognate students. Before the module there was a significant difference in performance of the two groups, but no significant difference between them after the module, suggesting that the skill gap had reduced. The main area in which scores had increased related to statistical knowledge, while there was no significant change to confidence levels or attitudes to quantitative skills.

Overall, in the formative assessments, there was no significant difference between the results of cognate and non-cognate students. Analysis of the results to the formative assessments was carried out for non-cognate students who had completed the preparatory module in semester 1 and those who had not. In general, those who had undertaken the preparatory module achieved slightly higher scores than those who had not, but as the number of students in the former group was small (N=11), those differences were not significant statistically.

The summative assessment results showed no significant difference between cognate or non-cognate students in terms of their data skills. Again, non-cognate students who had participated in the preparatory module scored slightly higher but the difference was not significant (again with a small sample N=11).

Feedback from students about the different approaches to teaching in the two blocks revealed that the non-cognate students liked and were more satisfied with the inquiry-based

and collaborative approach in the first block than with the second block, which used the flipped classroom and reflection approach.

Key learning

- Generally, the DATA module succeeded in increasing the data skills of non-cognate students and their level of confidence in using them. There was some evidence (but not statistically significant) that non-cognate students who participated in the optional preparatory module during induction performed somewhat better. Also, non-cognate students strongly preferred the peer-based collaborative learning approach in block 1 than the flipped classroom approach in block 2.

- Limitations were the modest take-up of the preparatory module during induction and its optional nature meaning that students did not complete all the possible activities. Making this module mandatory would increase participation in it, although this could lead to congestion in the timetable for semester 1, so another solution could be to add 1-2 weeks of intensive preparatory study prior to the start of semester 1.

- Despite the lack of statistical proof of the difference the preparatory study made to students' performance, due to the small sample sizes, Solent strongly believe that it is effective and will seek to continue to incorporate it in these and potentially other courses where there are non-cognate students.

3.2 Lancaster University

The Psychology Department at Lancaster University used the funding to deepen its evaluation of the ongoing change in statistical software used to teach undergraduates statistics for Psychology in Years 1 and 2, from SPSS to R. With most of the students having limited mathematics background, the explicit aim of the change is for students to develop competent knowledge and use of an open-source tool for statistical analysis that demands conscious choices for their data. Hypothetically, these choices can be traced in students' R scripts, heightening the transparency of any data analysis workflow, compared with other tools, as well as the potential reproducibility of research. Lancaster began the roll out of teaching statistics through R and RStudio in 2020/21, so this snapshot was taken midway through the planned changes.

RStudio is menu-driven, so creating, opening and saving files is easy for novice students. After that, R works by the execution of commands written by the user. This demands a level of data skills knowledge and use far beyond the making of drop-down menu choices in SPSS (or some other tools) and is a new and unexpected learning experience for students. The funding offered the opportunity also to evaluate whether and how Lancaster assesses the data skills learning component of the statistics curriculum.

Project activities

Several strands of research were conducted with the students, utilising data from weekly assessments, six scheduled class tests and focus groups, plus surveys and focus groups with staff. Broadly, most students performed well from the outset and, at the end of their

training, most Year 2 students felt confident in their basic skills in R and RStudio alongside their statistics learning.

Findings

The assessment strategy showed that there were no statistical differences between male and female students, nor between those with or without prior A-level maths or with or without prior A-level psychology. There appeared to be very little difference in the range of scores for the 2021 student cohort taught using SPSS and the 2022 cohort using R. This suggests that the two-year training programme offered a general learning experience to all, irrespective of their prior learning before entry to Year 1. The project did however raise the question of whether the assessment strategy, which has breadth of coverage for the taught curriculum, assesses the depth of learning or competency in generating independent data analyses sufficiently. Focusing on statistics knowledge and practice in the learning outcomes and assessments, it was felt that although use of R and RStudio demands that students learn a range of data skills, these are not currently highlighted, discussed or assessed as an object of knowledge or a set of discrete skills. There were fewer assessment questions related to coding or data skills, and those were less likely to be answered correctly, compared with questions focused upon statistical knowledge and application.

This was evidenced by many students signing up for focus groups to talk about their data skills and statistics learning but fewer about skills mapping (as if this was a concept unfamiliar to them). This suggests that data skills learning needs to be more explicitly framed in curriculum learning outcomes to ensure that teachers draw attention to it and that students understand this part of their learning has value and currency and develop the language to be able to talk about their level of competency.

Introduction of a skills-map self-assessment tool may raise the profile of data skills learning. This will now be embedded, initially at a small scale, in the curriculum and its value investigated over the next couple of years. A crude system of ‘traffic lighting’ may be used to indicate when students move from no use or awareness of a skill (red) to awareness (yellow) and finally to competent use (green), in their self-assessments. Ideally, this would be a process that students repeat periodically, alongside class tests.

Staff also discussed the possibility of project-based assessment rather than a multiple-choice quiz assessment. This could assess depth of understanding and cross-validate students’ self-assessments of data skills competency. An added value of project-based work would be a rehearsal of the workflow and processes that students will likely follow in their third-year project data analysis.

Surveys of teaching staff revealed they had very mixed knowledge of and experience with R and RStudio, and most were self-taught. Those with least experience felt that they were not confident about how they could supervise third-year students doing projects using R. Although it is not expected that all staff should use R and RStudio in the same way, a strategic plan is needed to upskill teachers in use of the software in order to increase capacity and resilience.

From the focus groups, students themselves did not envision working with data as part of their post-degree career, which they tended to see as focused on ‘helping people’. They

seemed to under-appreciate how the data skills and statistics learning would empower them to evaluate and produce evidence-based practice that would then help people. This seems to reveal a gap in provision, missing an opportunity to underline the importance of evidence-based practice in therapeutic settings and how statistics training enables the psychology graduate to conduct action research or small-scale studies which lead to changes to institutional practice based upon reliable data and findings.

Key learning

- Through this short project, Lancaster has found that its students are developing data skills almost as a by-product rather than as a focused endeavour, and that current assessment methods do not tap data skills learning.
- Currently, data skills are not assessed formally to the same extent as statistical knowledge or understanding, and where they are assessed the students perform less strongly. To capitalise on the learning that is happening, learning objectives for data skills need to be explicitly framed in curriculum learning outcomes.
- Students struggled to comprehend the utility of data skills for their future careers; further work is needed to explore with third-sector and industry partners how to best communicate to students the importance of data skills training, and the value of these skills for working in psychology careers. This could be in the form of curricular development, or through work placements.

3.3 Birmingham City University

The institution-wide Data Strategy Assessment project at Birmingham City University (BCU) aimed to understand how students experienced different approaches to teaching data skills, and identify effective teaching and learning practices for statistics and data skills education. It also explored statistics anxiety in the BCU student population. This project examined five non-cognate subjects, all of which have core teaching staff who teach statistics and data skills in unique and applied ways: BSc Psychology (and pathways), BSc Biomedical Sciences, BEng Biomedical Engineering, BSc Business and Finance and BSc Finance and Investment.

Data skills are taught in different ways across the courses, scaffolded though a framework which provides a foundation set of skills is embedded across Level 4, with successive modules at Levels 5 and 6 building on these to ensure graduates understand and can apply core data skills. Central to these teaching approaches are threshold concepts and active problem-based learning. Various software systems are used across the courses, chosen to meet graduate requirements for each field.

Project activities

Data was collected from a survey of students and from focus groups, and historic student progression data from across the five non-cognate courses was also analysed. There were 240 responses to the survey, and a further 51 students from across the five courses participated in the focus groups. Around 291 students in total were within scope of this research. In addition, three interviews with module teaching staff and an Associate Dean

affiliated with teaching and learning at BCU were also undertaken to investigate staff perceptions of student engagement, barriers to delivery, and effectiveness of current approaches.

Findings

Whilst, initially, the intention was to use learner analytics to understand any link between students' prior qualifications and their attainment on the course, as well as any differences relating to the approaches used to teach data skills, analysis of progression data indicated no statistically significant findings. In using mathematical scenarios within the student survey to assess students' confidence and understanding of basic data concepts, no evidence was found to suggest that students with weak GCSE Maths grades were any less confident than those with strong grades. In fact, on some occasions, the reverse was found. Instead, the analysis of data collected during this project focused on staff and student perceptions of data skills, and of the data skills teaching strategies employed at BCU. This particularly related to levels of students' relative confidence and anxiety in relation to learning data skills, and also whether prior attainment shaped these learning experiences.

When asked about their enjoyment of data skills, around a quarter of all students surveyed indicated that they did not enjoy any aspect of their learning on statistics and data skills modules. This only varied modestly by course, from 20% of Biomedical Science students to 30% of students from Psychology and related courses. This also varied slightly by gender, with 28% of female respondents but 22% of male giving this response, with a broadly similar difference on all courses.

Perceptions of the value of data skills also varied, though students were more likely to assess statistics and data skills as being useful for future employment than they were to enjoy learning them. Nearly half (46%) of all students felt that being taught statistics and data skills was relevant for their future studies or future employment, though over a third were unsure and 18% felt it was not. Again, perceptions differed by discipline; while in Psychology only a third of respondents felt learning statistics/data skills was relevant to their studies and/or future employment, more than half of respondents enrolled on the other four courses felt that it was relevant.

Whilst statistics anxiety was not specifically defined in this project, a number of measures related to this were included in the survey and focus group discussions. Findings highlighted that, prior to enrolment, the majority of students (69%) did not feel confident with their grasp of statistics and/or data skills, and those studying Psychology were far more likely to give this response than on other courses. Of those respondents who indicated that they were not confident, nearly two thirds (60%) were female, indicating the lower levels of confidence of female students with data skills and statistics. When asked whether the word 'statistics' made them feel anxious, again students on Psychology courses were more likely than their peers on other courses to indicate that they felt anxious. Analysis highlighted evidence of a link between prior attainment and statistics anxiety, with more of those with lower prior attainment expressing anxiety.

Findings from the student survey indicated that the majority of respondents found strategies to support those with less confidence and ability with data skills and statistics helpful, such

as provision of pre-course material and assessments, particularly in Years 1 and 2. Interviews with staff highlighted how it was felt that opportunities for students from different disciplinary backgrounds to learn data skills together could be beneficial, but that timetabling constraints were a barrier to this. Survey findings showed that the vast majority of students across all five courses had found additional support classes helpful if they had attended any (although nearly one third had not). Across all courses, as would be expected, those who most appreciated the additional support classes tended to have weaker prior grades.

Key learning

- The lack of specificity about the inclusion of data skills and statistics as components of these courses (aside from in Psychology, where it is mandated by the British Psychological Society and thus stated in pre-course information) led to many students failing to appreciate the volume and relevance of statistics and data skills for their chosen course.
- Significant proportions of students on these non-cognate courses did not consider data skills as relevant to their future studies or employment, indicating the need to explain better to the students the inclusion of these skills within their learning.
- Although there was no evidence to link prior attainment with competence in basic data skills, learners who had achieved lower GCSE Maths grades were more likely to express anxiety than their peers with higher grades.
- Female students were less likely to enjoy learning data skills and statistics, and were significantly less confident about their abilities, than male students.

3.4 University of Hull

The purpose of the project was to evaluate approaches to the delivery of data skills in non-cognate disciplines at the University of Hull and assess students' perceptions of their learning gains. This project aimed to assess stakeholders' perceptions of data skills teaching, and the perceived learning gain of students taught in different disciplines and using different methods. This project assessed the different approaches to the teaching of data skills across six subject areas at the University of Hull: Biology; Biomedical Sciences; Psychology; Economics; Criminology; and Geography. Four of the disciplinary areas included in this project have accrediting bodies that require students to acquire graduate level data literacy and data skills: Biological Sciences; Biomedical Science; Psychology; and Geography.

The predominant approach to the teaching of data skills at the university is through the provision of one or more stand-alone skills modules as part of the curriculum, delivered by teaching staff in the subject area, with opportunities for students to demonstrate higher extents of competence in data skills through research projects during their final year of study. This approach to data skills teaching – through stand-alone modules – was used in five of the above subject areas, whereas in Biology, data skills are instead embedded throughout the curriculum, and data skills are attached to specific learning and assessment tasks.

Project activities

Quantitative data were collected through an online survey distributed to students throughout the 2021/22 academic year. The survey collected data on degree course and level of study, current level of attainment and previous experience of data skills learning. However, engagement was low, with 61 respondents. Around 600 students were in scope for this project.

Qualitative data was collected via interviews with teaching staff and focus groups with students in spring 2022. Members of teaching staff were interviewed, with three from each of the six departments represented in the sample. Despite the use of incentives, recruiting students from across subject areas for the focus groups was more challenging, with 10 participants in total.

Findings

Some insight was gained into students' experiences of data skills teaching and their motivations in relation to learning data skills, as well as in understanding how successful teaching staff felt current approaches to data skills teaching were.

In examining students' perceptions and motivations in relation to their data skills learning, whilst many student respondents indicated that they were interested in, and open to, learning new data skills, they were much less likely to enjoy this learning. Further, whilst respondents appreciated the value of data skills to their degree programme, and perceived them as useful, particularly when undertaking a research project or dissertation, views were more complex in perceptions of the longer-term value of data skills. Whilst data skills were perceived by the majority of respondents as being helpful for future careers, nearly half did not see learning data skills as related to their personal goals.

Considering confidence with data skills, student respondents expressed higher levels of confidence in relation to collecting data, recognising patterns in data, and in data management, compared to lower confidence with using specific methodologies and software. Other challenges that students reported were the terminology required to learn data skills, which could feel overwhelming, particularly in scaffolded learning disciplines where modules were dedicated to data skills. Negative student experiences of data skills teaching could lead to disengagement and negative perceptions of the value of data skills, particularly when this learning was not required outside specific data skills modules.

The teaching staff interviewed felt that data skills and data literacy were vital for students' understanding of their discipline and their academic development. Staff also considered learning data skills to be useful in the longer-term, for postgraduate study, or for employment in a cognate research field. However, for staff across disciplines, it was perceived that lack of prior experience and lack of preparation for learning data skills led to low confidence and engagement from some students. Some mentioned that within their teaching, they aimed to address lower levels of competence with data skills and build students' confidence in numeracy and data literacy.

Some disciplinary analysis of data collected with teaching staff was undertaken, which highlighted how some in Biomedical Science and Criminology felt that students needed more

opportunities to apply data skills between being taught them in the first year and then applying them in their final year. In contrast, Biology staff indicated that though students often arrived with differing levels of ability and confidence with data skills, they had observed students make good progress and grow more confident, and felt that the embedding of skills across the programme reinforced learning and contributed to significant learning gains.

Key learning

- Whilst the second aim of the project – to assess the perceived learning gain before and after data skills teaching – was not possible due to low student engagement with the learning gain tool, it was felt that the survey questions were an effective means of assessing learning that could be utilised in future work.
- There were no differences identified in students' indicators of learning (as measured in the survey) between the various approaches to data skills teaching, though a judgment on the comparative effectiveness of teaching data skills through either discrete or embedded modules was not possible, due to low participation in data collection activities. However, the project team concluded that for five of the six subject areas, "*the stand-alone module model does not appear to support deeper application of data skills across the curriculum, but students do appear to develop the skills that they need to succeed in final year assessments, and staff do see these competences in data use*".
- The embedded approach to data skills teaching recently adopted by the department of Biology is new within the university and should be more completely evaluated. However, it was evident that this approach may necessitate increased support and training for teaching staff, who may have differing levels of experience and ability in relation to data skills.

3.5 University of Wolverhampton

The project at Wolverhampton involved the mapping of existing data skills teaching approaches across all courses at the university, as well as widening access to an existing online fundamental data skills module, currently compulsory for students on cognate courses in the Faculty of Science and Engineering. This part of the pilot project involved rolling out the module, which includes resources, short videos and quizzes for students to develop essential mathematics and programming skills and test their own knowledge, to all Level 5 students across non-cognate disciplines. Student competence and confidence with data skills was assessed both prior to and after undertaking this module.

As part of the project, research assistants conducted a review of all existing pedagogical approaches currently utilised to embed foundational data skills in non-cognate subjects in the university, mapping how and where data skills are taught. Paid interns were recruited from conversion MSc data science courses at the university to support data collection.

Project activities

The first strand of data collection was from 'pre' and 'post' surveys, built into the online data skills course, which was aimed at more than 1500 students. The first survey aimed to understand students' data skills abilities and their confidence prior to undertaking the

module, and the follow-up survey was to assess the effectiveness of the module in developing students' foundational data skills. In addition, semi-structured interviews with teaching staff were undertaken to understand perceptions of the most effective strategies for data skills teaching in their subject area. Interviews were also conducted with students, to better understand how they perceived their data skills teaching.

A final strand of data collection was to undertake a range of large-scale web-scraping exercises to understand both how data skills are taught across the entire university and to understand the currently validated provision of data skills across UK universities. This involved the extraction of data from module and course descriptors from over 110 UK HE institutions. The exercise generated a highly detailed dataset on the provision offered at all UK universities, although as analysis was still underway at the point of this report there are no findings to report yet from this strand of activity.

Findings

The online data skills course was still underway at the end-point of the project, and the 'post' survey yet to be run, so findings are limited. However, emerging findings indicate that students feel positive about this approach to embedding foundational data skills in their learning. Respondents expressed a strong preference for undertaking the data skills course at an early stage of their studies, before the start of formal assessments.

Two findings about course content emerged from students' feedback. The first was that students often struggled to comprehend specific language relating to the regulatory aspects of data science and the implications of data protection and the General Data Protection Regulation (GDPR). This is a pedagogical challenge in relation to engaging students with these wider issues and facilitating their understanding. A second was that attempting to engage students with software as part of the online course had limited success, with some students appearing to fully engage with the software and others focusing simply on completing the assessments posed within the course.

Interviews revealed further issues relating to terminology used to discuss data skills. For both staff and students there was perceived ambiguity in the distinction between fundamental/foundational data skills and fundamental/foundational mathematics skills. This indicates the need to try to establish clear and distinct definitions which are widely recognised and understood.

In terms of perceptions of approaches to data skills teaching, students felt that in some cases the way in which data skills was embedded into their course was of context with the rest of their learning. For example, existing modules were sometimes replaced with a module or unit on quantitative methods or statistics, which could make this feel disjointed. Early analysis indicates that this was more commonplace in traditional subject areas, for example the built environment and engineering. It appeared that staff in some subject areas were more willing to embrace creative and innovative methods. For example, in sports science and health studies it appeared that there was significant appetite to revise the curriculum to incorporate new and innovative approaches to data skills, including specialist software.

It is hoped that the online foundational data skills course will be extended, with plans to make it a mandatory requirement for all Level 5 students in their first semester.

Key learning

- Bringing MSc students into the project by employing them as interns was valuable in developing their ability to work on a specific data-based project and engaging them in their learning – the datasets generated will also be used as a teaching tool in future courses.
- There is evidence that the terminology and language around data skills was challenging for non-cognate students and, for teaching staff too, the lack of specificity in definitions of foundational/fundamental data and mathematical skills was not helpful.
- There are additional challenges in engaging students in regulatory aspects of data skills, particularly in relation to GDPR and data protection.

3.6 Teesside University

The aim of the Teesside project was to understand the different approaches to data skills currently being delivered in non-cognate subjects at the university, and highlight areas where improvements could be made. The delivery of data skills training was perceived to be variable across the institution, with some departments such as education being supported through 'service teaching' where staff from other departments are used to support delivery of data skills teaching, while others rely on their own staff to do so.

The project focused on three modules which specifically teach data skills; one a core module for the science foundation year and the other two being undertaken in Level 4 of programmes in accounting/finance, cybersecurity, business and IT. Largely, students applied the skills they learned in these modules later in their course within projects and dissertations. Around 400 students were in scope for this project.

The main objectives were to establish a detailed picture of how data skills are taught on these programmes, understand staff and student perceptions of this teaching, and assess the effectiveness of these various approaches. It was also considered important to consult industry partners about their expectations of graduates in relation to data skills, in order to assess the fitness for purpose of current approaches to teaching.

Project activities

Data was collected from module evaluations, a systematic review of teaching and assessment documentation, a student survey and a student focus group. In addition, small-scale surveys of teaching staff and industry partners were undertaken to understand their perceptions. A small focus group with industry representatives was also conducted, following the survey.

The data collected from students was focused on those studying the data skills module that is a core part of the science foundation year. There were 40 responses in total to the survey, a response rate of 14%, while four students participated in the focus group.

Findings

Responses to the student survey highlighted that the majority of those who had completed the foundation module were satisfied with the teaching, learning materials, organisation and assessment. However, only around half of respondents felt that the module had increased their interest in data science or confidence using the techniques and principles in relation to big data and ethics. Students indicated they wanted more relevant context for their learning so it was more relevant to their specific course, as well as more focus on using specific programmes that would support them to collect and visualise data.

Student evaluation data for the three data skills modules within scope indicated some areas that can be improved. Two out of the three modules had pass rates that were lower than expected, indicating that current teaching and learning methods needed to adjust in order to support more effective learning. Feedback from students in the focus group indicated that in the foundation module there was an over-reliance on teaching how to use Excel, when they often needed specific knowledge of programmes such as R and Python later in their studies. Formative feedback, provided during the module, was particularly valuable in support of students' confidence and capabilities in using data skills.

Whilst digital citizenship and data skills form some of the expectations of the university's Academic Enhancement Framework, analysis of module content revealed that the curriculum and approach to teaching data skills in non-cognate subjects vary. Data from academic staff indicated that they perceived the most important data skills to be critical thinking, writing reports based on data, problem solving and research. Staff felt that, generally, where data skills were linked to real-world problems and immersive scenarios, this was most effective in engaging students.

Industry partners provided some insight into the types of data skills they valued most and identified graduate skills gaps. They indicated that the types of data skills most often utilised in their work were use of spreadsheets, data visualisation, writing reports and problem solving. They felt graduates needed to have more knowledge of database tools and machine learning. Significantly, there was evidence of some mismatch in how academics and industry partners valued certain aspects of data skills learning. Whilst industry representatives highlighted spreadsheet, data visualisation and database tools as very important skills, academics did not identify these as some of the most important. These insights will be used to inform curriculum development for teaching data skills to non-cognate students.

Key learning

This project has identified some key learning which can be used to inform and improve data skills teaching at the university.

- A solid grounding in data principles and skills was seen as necessary at an early stage, before students were asked to undertake final projects and dissertations, especially where these involved working with industry on real-world projects.
- The most effective methods for data skills teaching were felt to be those which encouraged students to work collaboratively, and in an immersive way with data. The

use of case studies which enabled students to understand how data skills apply to real world situations were perceived to particularly support engagement.

- Consulting industry partners about the most valuable aspects of data skills, and understanding perceptions of graduate skills gaps, was helpful in curriculum development.
- A consistent approach to data skills teaching was recognised as something that would be valuable to develop, potentially through establishing a common foundational data skills curriculum shared across the institution. This would enable teaching staff to develop a shared understanding of data skills pedagogy and ensure consistent student experience.

3.7 University of Newcastle upon Tyne

The project undertaken at Newcastle broadly aimed to develop and deliver training for those delivering foundational data skills teaching across disciplines. This initiative reviewed provision at the university and four other universities in NE England and involved work with charitable partners and employers to build a community of practice around foundational data skills teaching.

The project responded to challenges with the 'service teaching' model, a common approach to data skills training in non-cognate disciplines where teaching may be delivered by data science or computing specialists from another department. The project team believed this model might limit the capacity of teaching staff in disciplines to develop their curriculum in the most effective way for students. A key aim of the project was to upskill and build the confidence of non-cognate teaching staff through training and development. A community-of-practice approach and a 'train the trainer' model were used, alongside a mentoring programme, to address these issues and help build a self-sustaining community.

Project activities

The project involved four strands of activity:

- Evaluation of existing provision, including surveys to understand challenges facing teaching staff and student perceptions of data skills and their teaching.
- Development of training materials, co-created with the involvement of students and teaching staff.
- Training of those teaching data skills to non-cognate students, supported by mentoring.
- Industry and public sector engagement – in total, individuals from 52 organisations took part in engagement in the project.

Findings

38 of the targeted 50 non-cognate teaching staff, from across three faculties, were trained in teaching data skills. A short guide was produced to help potential participants achieve buy-in to attend the training, by articulating the positive impact that it could have on their teaching.

Interviews with non-cognate teaching staff indicated a range of challenges for those involved in data skills provision, including some mismatch between what employers need and what is currently taught, and insufficient focus on the applicability of the skills to real-world scenarios.

Developing appropriate resources and curriculum content for diverse learners with different needs and prior experiences of data skills was seen as a key challenge. Significantly, the lack of enough appropriately qualified teaching staff to deliver high-quality data skills teaching was highlighted. The need was highlighted for initiatives such as the 'train the trainer' scheme to build capacity. Students' anxiety in relation to statistics and data skills was also seen as a challenge.

Consultation with industry representatives via a survey and focus groups highlighted three main findings. First, there was perceived to be a lack of understanding by academic teaching staff in relation to the data skills which are practically needed in industry. Second, industry partners felt there could be improved routes for knowledge exchange and improved connections with instructors delivering data skills teaching. Finally, industry representatives indicated that a key barrier to collaboration with universities was the disconnect between institutions and resulting inconsistency between different graduates and their data skills.

Findings from the project have been leveraged to deliver an extra-curricular training intervention in collaboration with Analyst Network North-East and the National Innovation Centre for Data. Over 100 individuals from the five universities participated. This provided data skills training and an opportunity to work with North-East England Climate Coalition, a cross-sector initiative. A further £25,000 of funding was secured from the Engineering and Physical Sciences Research Council to extend the work undertaken in this pilot nationally, via a large-scale survey of 1,000 adult learners and 1,000 business owners/proprietors. This extension to the work will seek to explore regional and sectoral trends in the identified themes at a national level. The project has also had impact via contributions to The Alan Turing Institute's Data Science and AI Educators' Programme to share lessons learned to support upskilling non-cognate teaching staff to deliver data skills teaching.

Key learning

- There is evidence of a mismatch between the focus of existing data skills provision and what is needed by industry, which appears to represent broader issues relating to the need for improved routes for knowledge exchange.
- Those involved in delivering data skills teaching highlight the need for capacity building across disciplines, so that more staff in non-cognate disciplines are appropriately qualified, trained and confident in delivering high-quality data skills teaching to students.
- This project brought non-cognate teaching staff together to share best practices and gain confidence in teaching data skills, and there are plans to engage more staff yet. A mentoring process was also set up to support participants in their skills development.
- Newcastle are hosting a two-day Data Skills Education 'sandpit' in September 2022 to support the update of the lessons learned from the project, intended to attract participation from industry, academic and policy organisations nationally, and provide avenues for cross-institutional collaboration.

4 Observations and lessons learned

4.1 Overall findings

Here we draw out some high-level findings and extract key learning from the evidence review and pilot projects.

Data skills

- Reflecting wider issues around appropriate terminology and definitions of data skills, there continues to be some lack of clarity about the scope and language of ‘data skills’ (as opposed to broader ‘data literacy’, for example) and competencies, and the extent to which these should be considered ‘foundational’ in HE institutions.
- There was evidence from some of the pilot projects of some misalignment between the skills that industry states it seeks, what is being taught to students currently at foundational level and the skills that students require to undertake project or dissertation work in later years of their programme.
- Pilot projects reported that a number of students were not acknowledging the value of data skills within their own potential career trajectories or life contexts, suggesting that there is scope for more work contextualising this learning and demonstrating the relevance of these skills to a range of future careers.
- There was some evidence of ‘statistics anxiety’ (and similar anxieties) amongst non-cognate students, and that some teaching staff find non-cognate students' varying levels of data skills experience/ability to be pedagogically challenging.
- These issues could to some extent reflect the relative immaturity of data skills as a ‘discipline’ and also their wide applicability across many non-cognate subjects. In contrast, there is a longer history of teaching statistics in subjects such as psychology, for example, from which it may be possible to learn.

Approaches and effectiveness

- We were able to identify quite a wide range of approaches to data skills teaching, or supporting students to learn them, ranging from embedded content or mandatory modules within the subject curriculum taught by subject team, right through to individuals being encouraged to undertake self-study using third party resources such as MOOCs. There seems, however, to be little published evidence about the effectiveness of most of these approaches specifically in relation to data skills.
- The wider evidence review suggested that relatively few undergraduates take up their university’s own data skills programmes where offered outside their subject curriculum, or external self-study options, presumably because at that level their preference is for data skills to be taught within their curriculum.
- One of the pilot projects found some evidence for non-cognate students to prefer interactive approaches to teaching rather than those based on self-study (or a ‘flipped classroom’ model).

- A number of pilot projects found some evidence that modules embedded within the subject curriculum, featuring examples located within the subject, were more popular with students than standalone modules or content provided in a service teaching model. 'Real world' applications and examples, relevant to the non-cognate subject context, appealed most to these students.
- There was a general perception from across the evidence that tying data skills learning into wider employability development could be effective.
- Pilot projects reported perceptions that the earlier within a course that data skills are taught the better – giving time to embed these skills and demonstrate their importance to students well ahead of when they will need to use them in final projects/dissertations.
- There was a little evidence (albeit not statistically significant because of the small sample size) that an optional module introducing data skills early in the course could increase attainment in a subsequent mandatory module, although this begs the question of how to encourage students very new to a course to undertake such additional study without putting them off.

Implications

- There is some emerging evidence that approaches to data skills learning that are embedded within a non-cognate subject curriculum, taught by the subject team, seem to work best for undergraduates. This has implications for the upskilling of subject teaching staff (and could be more resource-intensive than a service teaching model).
- There may be an opportunity for a national staff upskilling programme such as Q-Step to facilitate that additional learning by teaching staff.
- There seems to remain a lack of evaluative evidence on the effectiveness and impact of different approaches to teaching data skills, and the extent to which these short, modest pilot projects were able to develop such evidence was limited.
- Several projects have been able to start to scale up their activities and translate their emerging findings into other contexts, in some cases pulling down other funding to pursue upskilling of staff, while others have embedded particular models into a wider range of courses as a result of the pilot.

Other observations

- There was a relatively high level of interest from HE providers to address the issues considered in this project and a desire to do more work in this area. Several pilot project teams indicated that they would continue to attempt to develop and evaluate their approaches beyond the funding period.
- The short timescale for these pilot projects, and relatively small amount of funding available to each, meant that the amount of evaluative work that projects could achieve in practice was limited. Many projects struggled to engage large numbers of learners in the research within the timescale available, limiting the investigations possible.
- One pilot project's utilisation of students from its MSc data science cohort to undertake evaluative studies about data skills development was innovative and effective.

4.2 Recommendations

1. DCMS should work with key stakeholders to develop and agree appropriate terminology and definitions for data skills and data literacy, including the extent to which these should be considered ‘foundational’ in the context of HE provision.
2. The sector should undertake further strategic consultation with industry to establish the types of data skills that employers are seeking from graduates, to ensure that curricula and other data skills learning is as fit-for-purpose as realistically possible.
3. The sector should consider a focused programme of work to develop a range of materials demonstrating the value of data skills to those working in a wide range of early careers, so that more students appreciate their potential future value.
4. Providers need to undertake longer-term, substantive evaluation of different approaches to data skills teaching, building on the pilot projects in this study, in order to obtain a more robust evidence base about the effectiveness of approaches.
5. The sector should consider the potential value of a national programme to upskill non-cognate HE teaching staff to build their confidence in delivering data skills teaching, highlight best practice in teaching students with mixed abilities and prior experiences, and share practice and resources.

Glossary of abbreviations

AI	Artificial intelligence
CPD	Continuous professional development
CRAC	Careers Research & Advisory Centre
DCMS	Department for Digital, Culture, Media and Sport
EDUCE	Experiential Data science for Undergraduate Cross-Disciplinary Education (at the University of British Columbia)
EdX	A MOOC platform combining many universities' offerings
GCSE	General Certificate of Secondary Education
GDPR	General Data Protection Regulation
HE	Higher education
Jisc	Joint Information Services Committee
MOOC	Massive Open Online Course
NDS	National Data Strategy
OfS	Office for Students
R	A software language and environment for statistical computing and graphics
SAS	A computer programming language used for statistical analysis
SQL	Structured Query Language
SPSS	A statistical software suite developed by IBM
VLE	Virtual learning environment