A flexible school and college level qualification in Data Science

Kate Farrell
kate.farrell@ed.ac.uk

Jo Watts
jo.watts@effini.com

Judy Robertson
judy.robertson@ed.ac.uk

Tommy Lawson
tommy.lawson@ed.ac.uk

June 20, 2023

Abstract

This paper describes the design and development of an innovative high school-level qualification in data science. The qualification has been available for 4 years; 1982 learners have completed the course to date across 30 educational institutions. We describe the structure of the course, its pedagogical principles, and initial feedback from teachers and learners. Based on our experiences as members of the qualification development team, we discuss the prerequisites to successful adoption by schools and colleges. We conclude with a set of recommendations for other educators who are designing similar qualifications in other school systems.

1 Introduction

Living and working as active citizens in an increasingly digital world requires data literacy—a core set of practical, critical and creative skills necessary to investigate data sets to analyse real-world problems. There is increasing demand for university and professional courses in data science to meet the need for data skills in the employment market. Consequently, there is a requirement to give high school learners the opportunity to develop data literacy and data science skills to equip them for life and work. School-level data science curricula are beginning to emerge: Mike et al. (2020) report a curriculum for high school computer science learners (15 year olds) in Israel which explores "what data is, how it is collected, and how it can be processed to solve real life questions and problems" while balancing this with in-depth knowledge of machine learning algorithms. By contrast, in their draft data science curriculum for German secondary schools (targeting 15-18 year olds), Heinemann et al. (2018) identify the importance of teaching learners about the societal implications of data science, rather than focusing only on technical aspects. This paper describes the development of a national qualification in Data Science for learners in schools

and further education colleges. It describes the development process, key design principles and themes, reflects on what has been learned and concludes with a set of recommendations for the development of similar qualifications in other education systems.

2 Qualification background

In 2018 the Scottish Qualification Authority (SQA), the national accreditation and awards body for Scotland, commenced the development of qualifications in data science aimed at school and college learners. As an awarding body they develop a framework of qualifications for schools, further education colleges, employers and training providers. In the senior phase of secondary school, learners have the option to choose which national qualifications they wish to study, assuming that the school offers the course. The SQA also have a range of vocational qualifications "developed with industry partners, and responding to perceived skills needs." These include National Progression Awards (NPAs) which are aimed at assessing a defined set of skills and knowledge in specialist vocational areas. NPAs can be offered by either schools or further education colleges. The NPA courses normally consist of three units, each of which requires around 40 hours of learning time. Candidates for NPA qualifications demonstrate their knowledge and understanding through internal assessments throughout the course. Practical skills are assessed by performing a practical assessment rather than by using a written final exam.

This paper describes experiences from developing the NPAs in Data Science which are available at three levels, SCQF levels 4, 5 and 6. This is equivalent to Levels 2, 3 and 4 in the European Qualifications Framework. Learners in secondary schools typically study for qualifications at these SCQF levels between 16 and 18 years old. Adult learners of any age may study at these levels at a further education college.

2.1 Development process

When developing new qualifications, the SQA appoint a team with representation from industry, schools, colleges, universities, training centres, special interest groups and the public sector. The first author was the lead developer with the responsibility to carry out a feasibility study and guide the qualification development team throughout the process. The second and fourth authors were also members of the team.

The feasibility study aimed to solicit views on a new qualification in Data Science from different sectors of education and industry including content, assessment and potential barriers to delivery. Individual in-person interviews were conducted with seven people from a range of secondary schools, further education, local authorities, and industry. In addition, there were 19 responses from an online survey that was distributed using the SQA's communication channels. Overall, the response from all sectors was positive. Most respondents supported

the proposed qualification, with only a few answers unsure or stating a preference for the development of an exam-based National Qualification. The teachers identified a need for support and professional development to help them master new data science material.

From these findings and discussions, the structure and framework of the NPA qualification, including the individual units which would make up the award, were developed by the first author and then approved by the qualification development team. The framework specified both outlines for the creation of new course units and recommendations for the inclusion of suitable units from the existing SQA course catalogue.

Writers and Vetters were commisioned to write the individual units at each level, but due to the limited time available, they were largely writing in isolation from the rest of the team. Unit writers were either teachers or retired teachers. Each author wrote all the units on the same topic at each level (4, 5, 6), and the units were vetted by another person, generally a teacher or an expert from industry or university. The SQA team then approved or adapted those units before publishing them.

An independent group of experts from universities, colleges, schools, national organisations and employers formed a validation panel, which met in May 2019 and approved the qualification specification.

3 Design Principles

The NPA in Data Science is based on a set of design principles which are suitable for the context of this type of qualification within the Scottish education system. These principles may contrast with other school-level data science courses which have been developed in other countries depending on the purpose. As noted by Black and Wiliam, "in each country assessment practices have impacts on teaching and learning that may be strongly amplified or attenuated by the national context." (Black and Wiliam, 2005, p. 260). For example, in other contexts, a data science course may be compulsory for all learners, or embedded within computer science or mathematics classes, or require assessment through high-stakes examination. However, we hope that educators and policymakers in other educational systems will find it helpful to interpret and adapt the principles we document here.

At the time that the qualification was developed, there were very few existing curricular frameworks and those which had already been developed (such as the IDSSP framework or the Bootstrap Data Science course) approached the subject from the perspective of a single discipline (e.g., statistics or computing science). The IDSSP framework, although useful, was too complex, particularly for the Levels 4 and 5 of the NPA. Other frameworks and courses were aimed at undergraduate students or professionals with existing coding experience.

3.1 Continous summative assessment of learning

The main constraint the SQA had specified was that the data science qualification was to be a National Progression Award, which must be *continuously summatively assessed by coursework* rather than by the traditional exams which are more commonly used to assess the other national qualifications.

Learners are assessed by their teacher in class on their knowledge and understanding of the theory sections of the units, and they carry out a series of practical tasks to assess the skills component of the units (product-based assessments). These assessments can happen throughout the year whenever the learner is ready to demonstrate their competence. Performance-based assessments, including product-based assessments, are considered to be "one of the best forms of assessment" (Dixon and Worrell, 2016, p. 157) because they measure the application of knowledge rather than rote memorisation of facts, and can take into account the integration of knowledge across topics and evaluate decision making. Continuous summative assessment is an appropriate assessment method for the in-depth exploration of how a learner applies skills in their local context over time (Muskin et al., 2017) and is suitable for a diversity of learners (Kapambwe, 2010). Continuous assessment can be used effectively in complementary ways for both formative and summative purposes (Dixon and Worrell, 2016). In the case of the NPA, the teacher can fluidly and iteratively move between formative assessment of the learners' understanding of the data science units, to summative assessment of the skills they have mastered, as well as using information from the summative assessment to adapt teaching to the learners' needs.

Another advantage of continuous assessment is that it does not rely on high-stakes examinations which can induce anxiety and negatively impact the well-being of learners, particularly among girls, members of ethnic minority groups and learners who have disabilities (Von der Embse et al., 2013).

3.2 Interdisciplinary delivery and flexible options

The data science qualification was deliberately planned so that it could be *delivered by teachers across a spectrum of subject areas.* This is in contrast to previous school-level data science courses which have been embedded in computer science classes (Mike et al., 2020; Heinemann et al., 2018). However, this choice was appropriate to the context because previous research found that learning outcomes relating to data literacy can be found across many subjects within the Scottish curriculum (Farrell and Robertson, 2019), and so teachers of computing, maths, science, English and the social sciences would all be familiar with some (but not all) aspects of data literacy. This is a pragmatically important consideration because there is a shortage of computing teachers in Scotland (Robertson, 2019) and relying on only the computing teacher workforce to deliver the new qualification would be a barrier to it being an option for all learners. The NPA can be taught using the framework of data science but with a context of a specific subject area, such as a geography teacher using

case studies and datasets from satellite and land use data. This allows for more flexible timetabling based on staff availability, expertise, and interests. Commitment to the interdisciplinary design was reinforced by consultation during the feasibility study. Respondents to the feasibility study survey were shown three possible curricular models. The most popular model, which is the one that was then adopted, was favoured by those who liked the flexibility of choosing optional units depending on the staff's expertise and ability to teach the award over multiple departments. One survey response stated "the optional units allows schools to make a course offer that plays to the strengths of its staff."

3.3 Freedom to choose suitable software tools

The NPA has been designed to provide a great deal of *flexibility in the software tools* that can be used by learners for data gathering, analysis and visualisation. This is based on the principle that data science requires a set of intellectual skills and competencies which can be demonstrated using a range of different technical tools, and it should not be seen as a practical course in learning how to use a particular software package. Furthermore, the flexibility of software choice is also logistically helpful for delivery given the variety of software licensing arrangements and existing teacher knowledge across different regions of the country. In the feasibility study, one educator suggested that "another barrier might be access to appropriate software - so industry could help by allowing free access to platforms/tools for schools." It was very important to the development team that every outcome could be achieved using a selection of different tools that are free (or free for educational use) or open source and that schools should not have to invest money in software or hardware. It was also important that there were a range of tools that could be installed as well as a selection of online tools available, as some schools have strict restrictions on installing software.

In addition, the NPA can be achieved using only a basic spreadsheet tool. It is important that learners gain a secure grounding in their knowledge of data science and statistical concepts. Gaining experience of particular tools or programming languages that might be favoured by industry is much less important to learners at this stage of their career.

3.4 Other design considerations

Each of the units was written so that it could be *delivered stand-alone*. For example, the Data Citizenship unit could be delivered as part of a digital literacy course or an additional unit in a social studies course. This is advantageous because it increases the number of learners who will be exposed to data literacy or data science concepts as they learn about other subjects. The qualification was structured to *repurpose existing units* offered by the SQA. This enables learners to apply knowledge across qualifications and eases the burden for teachers as they familiarise themselves with teaching a new qualification.

The core units and most of the optional units have been written to *aid educators who are teaching multiple levels of learners in one class.* This is important to Further Education colleges, as finding a suitable time for Returners to Work and Upskilling students to meet can be challenging, so the ability to teach learners of different levels at the same time is beneficial. It is also helpful in schools where there is a shortage of specialist computing teachers.

4 Qualification structure and content

The NPA in Data Science consists of two core units at every level: a core unit in Data Citizenship and a core unit in Data Science. At Levels 5 and 6 learners undertake an optional unit as well as the two core units. The Level 4 course only consists of the two core units. This has been designed to allow learners the time to secure their learning at this level.

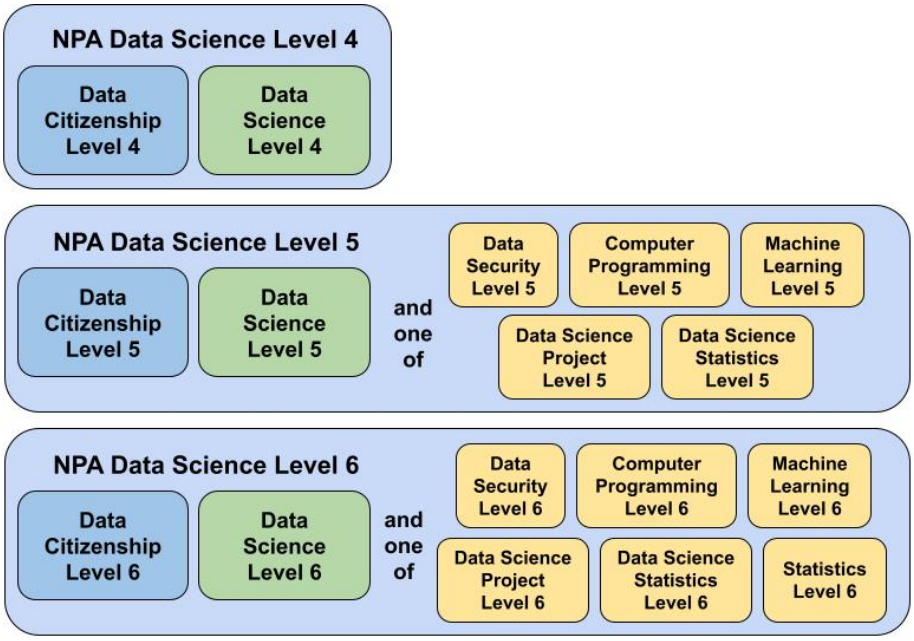

Figure 1: The structure of the NPA Data Science qualification at different levels.

New units were developed for the Data Science and Data Citizenship core units at each level. and for 'Data Science Project', 'Data Science Statistics' and 'Machine Learning', at Levels 5 and 6. Existing units from the SQA catalogue were adopted as optional units. These were 'Computing Programming' levels 5 and 6, 'Data Security' levels 5 and 6, and 'Statistics' at Level 6.

4.1 Overarching Themes

Three overarching themes that were established by the qualification design team as important across all the units in the course. It was important that *learners should have a secure understanding of the underlying concepts involved in data science as opposed to getting training in using a particular software tool.*

The topics of privacy and security were picked up by the feasibility study respondents as topics that had been missed in the proposal. It was felt that this should be a theme across the units and course. *Learners should ensure that they stay safe in all the work they do, they should keep their personal data safe, the data they use and gather in the course of their work should be kept secure.* They need to be aware of their legal and moral rights and responsibilities. Members of the qualification development team were keen for non-technical skills to be included. One member suggested "the '4 C's' of Communication, Content, Context and Compelling stories" and another member added "compassion". There was a strong agreement that the whole course should have a data for social good theme. It was felt that *learners should be encouraged to promote social good and social justice in the work they undertake in the course.* It was felt that this would appeal to different learners than would normally study computing subjects. This compassionate aspect to data science should include ethics. Across all the units the team wanted to encourage *learners should think ethically and morally in their work, question the implications of the technology on everyone involved in the process, and question any bias that may exist in the data or the algorithms.*

4.2 Core units

The Data Citizenship core unit involves understanding how data is used. Learners learn data literacy and basic statistics. They interpret data in different formats to find out interesting things from the data, investigate why unusual results or trends occur, and think about the impact or behaviour change resulting from the analysis. Learners also investigate how data can have both a positive and negative effect on society. The Data Science core unit involves learners gathering data from different sources then analysing it by exploring, modelling and validating the data. Learners then visualise the results and present on their findings, reporting on what they have found and how it can make a difference to themselves or others.

5 Experiences and reflections so far

5.1 Uptake in schools and colleges

The qualification has been offered for four years (although two of those years were subject to significant educational disruption as a result of the Covid-19 pandemic) and will be formally reviewed by the SQA next year. The number of

candidates who have entered the NPA qualifications at different levels, at school or college since the award was introduced in 2019 is shown in Figure 3.

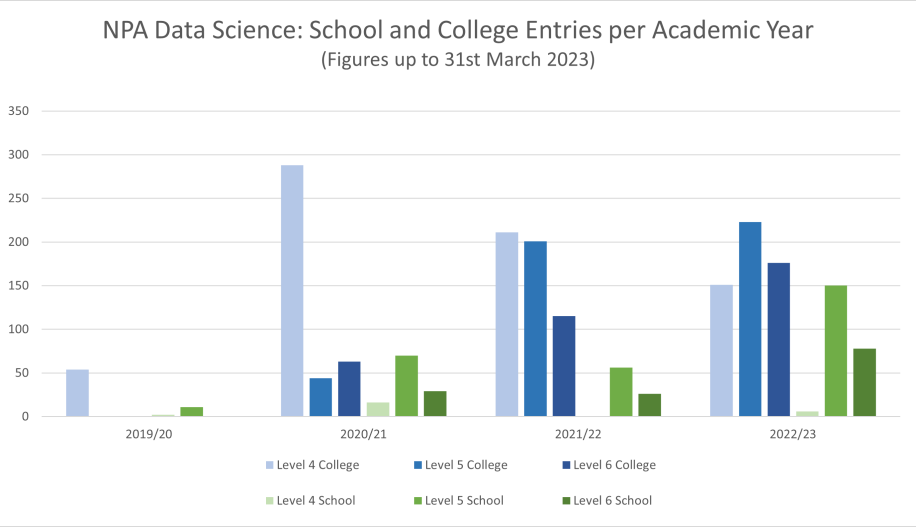

Figure 2: Entries to the NPA Data Science since its introduction.

The NPA has been more popular as an option for colleges than schools. 77% of the qualifications awarded since it went live in 2019 were to learners in colleges. It has been most popular as an additional numeracy qualification alongside existing college courses such as motor vehicle maintenance, beauty therapy or hospitality. Many colleges are also including the core units within other courses, in particular the Data Citizenship units. Up to the end of March 2023, 1982 learners had completed an NPA in Data Science, yet 3378 learners have gained a unit in Data Citizenship. This seems to be most popular at levels 4 and 5, with 463 learners gaining the unit at Level 4, 406 at Level 5, and just 212 at Level 6 in the most recent complete academic year, 2021-22.

5.2 Consultation findings

In 2021, the fourth author was commissioned to investigate barriers to delivering the qualification in schools and colleges, taking into account the relatively low numbers of entrants and awards. Evidence was gathered from surveys with learners and parents, interviews with teachers, college lecturers, and members of the data science industry.

The report focussed on the feasibility of using blended and online learning to increase the number of learners who could study for the qualification. It argues that the shortage of teachers with the skills to teach Data Science could be addressed by developing an online course, but that for this to work well various barriers to successful online learning which were encountered during the

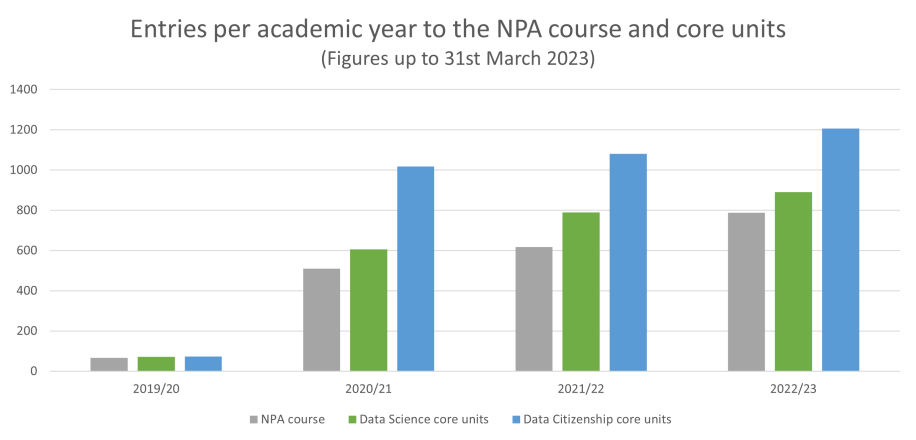

Figure 3: Entries to the core units of the NPA Data Science per year.

pandemic should be considered. It recommends increased professional learning opportunities and support for teachers, the development of additional educational resources to be used to teach the qualification, better online environments for practical work, solutions for remote assessment and further consideration of digital inequalities.

5.2.1 Relevance of the qualification for future careers

The school and college staff who were interviewed regarded the Data Science NPA as an up-to-date, modern qualification which will prepare young people for the emerging job opportunities in the sector and enable them to progress to study the subject in further or higher education. The teachers reported that employers are increasingly interested in learners' skills in these areas.

School staff commented that around half of learners had previously studied computing subjects and wanted to add to their qualifications in the subject. Of those without a computing background, many had chosen the course because they believed data science would strengthen their post-school options, both for higher education and future employment.

The consultation flagged that the perceived status of NPAs may be an issue for students, employers and school managers. Some schools report that NPAs do not enjoy the same parity of esteem as other longer-established qualifications when students are making course choices. This may be partly because NPA qualifications do not carry so many credit points during the university admissions process in the country. Data Science is also in a competitive curriculum area, often vying with computing and mathematics courses, which may restrict uptake. Work needs to be done to address these issues and to raise awareness of the progression routes and vocational opportunities offered by the Data Science NPA.

5.2.2 Additional resources and support for teachers

School staff who were interviewed during the consultation were dissatisfied with the available support materials which they required to teach the qualification. Some felt that the centrally provided course materials were too difficult for the level of learners. The teachers had to find and adapt materials and datasets to meet the qualification specification, which was time-consuming and burdensome. The teachers would like curated example datasets to be available for each level of study to save them time in locating and cleaning the datasets which are relevant, motivating, and sufficiently complex to provide contexts for students to apply their analysis, visualisation and interpretation skills.

The findings of this report are consistent with the common consensus from the feasibility study where teachers identified the need for support in learning the new subject matter. Their capacity for preparing for a new subject will be limited due to the time constraints on teachers currently. One respondent to the feasibility study said that "probably the greatest barrier will be staff competence in data science, and their confidence in delivering a qualification that crosses traditional school faculty structures (i.e. maths, stats and computing)." Another highlighted the need for a full set of teaching materials, wanting a "full course, that you can pick up and teach."

5.2.3 Engaging learners

The consultation identified that some learners had little understanding what was involved in the new subject or how it would benefit their future careers. In some cases, alternative computing topics such as cyber security sounded more exciting to learners and timetabling constraints precluded them from studying both.

Teachers reported low levels of student engagement for some of the theory elements, which they considered to be boring and lacking context and relevance for students. They found it challenging to create meaningful and interesting activities for these course elements, given the other constraints on their time.

Of the theory elements, students were most engaged in the issue-based topics of Data Citizenship, which could explore ethical issues. Students particularly enjoyed topics such as fake news, data bias, ethical use of data and why large commercial companies generate data.

Teachers and college lecturers reported that practical tasks were much more motivating for students than the theoretical section, were more challenging and required problem-solving skills.

5.3 Reflections on the qualification development process

One difficulty in developing a qualification in data science for school learners was that this was unfamiliar territory, particularly when the development process started. The SQA development process is perhaps better suited to developing courses in well-established subject areas. There was very limited scope or time for iterative development of an entirely new subject.

Due to this subject being unfamiliar to many teachers, it was challenging to take on unit writers who had expertise in data science. The background of the writing team was mainly computing science or mathematics teaching.

It has resulted in a lack of clarity on terminology or depth of knowledge required within some of the unit specification documents.

The lack of time within the existing SQA processes also meant that there was limited time to ensure consistency and reduce overlap across the different units.

It is difficult to recommend that all writers and vetters of units have subject expertise at the relevant level when these expertise is so rare. It is instead recommended that an extended development timescale for the qualification development team, an iterative writing process, and repeated consultation opportunities be factored into the development process of qualifications. Ideally the qualification and units should be reviewed after the first cohort has completed the course and learner and teacher feedback has been received. In addition, for qualifications in fast-moving technological fields such as data science and AI, there should be a shorter review cycle to ensure qualifications are up to date.

6 Ongoing work

Given the feedback from the consultation report, it is a worthwhile investment to support teachers through materials development for this new, highly interdisciplinary course. The course was designed to be taught by teachers from a range of subjects, which means that different teachers will need to upskill in different aspects.

The first and third authors led the development of educational materials to guide teachers and learners through each new unit of NPA in Data Science. This was considered particularly important because the qualification and the terminology of data science were new to teachers, although they would be familiar with some aspects of it from other curricular areas. The online materials consist of Learners' and Educators' Guides, full lessons, presentations, diagrams, practical exercises and cleaned example datasets. They were iteratively developed in consultation with teachers and learners who were piloting the new qualification. A four-day professional learning course for teachers was also offered.

To address the issue identified in the consultation regarding learner' lack of knowledge about the topic or why it would be beneficial to study it, the authors are working with a video production company to develop a promotional video to appeal to learners in schools, colleges and in the workplace to persuade them of the benefits of studying data science. It is hoped that schools can show this to learners at course choice time.

7 Recommendations

Based on our experiences of developing a national qualification in data science for school and college learners, we recommend that a data science qualification for school learners should:

- be developed by a team consisting of subject specialists as well as teachers with enough time for iterative piloting, learners consultation and refinement;
- undergo regular revision and refinement to reflect the rapidly changing nature of the discipline;
- be cross-disciplinary to enable teachers from a range of subjects to offer it to their learners (not just computing teachers);
- contain not only technical content such as programming or statistics but should emphasise real-world applications and societal implications;
- contain up-to-date and engaging topics such as current news stories, examples of 'bad graphs', political themes, running surveys and creating graphs;
- provide example datasets which are cleaned and pitched at an appropriate level to avoid confusing or demoralising learners with complex datasets.

In order to support teachers in teaching the new qualification:

- Local or national education departments should provide high-quality professional learning opportunities. These could be supported by subject-specialist academics;
- Comprehensive, accessible teachers' guides and other educational resources should be available to help teachers during the initial steep learning curve.

Acknowledgements

The authors would like to thank colleagues at the SQA and the teachers and learners who gave feedback on the qualification for their help and support.

References

Paul Black and Dylan Wiliam. 2005. Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal* 16, 2 (2005), 249–261.

F. Dixon and D. Worrell. 2016. Formative and Summative Assessment in the Classroom. *Theory Into Practice* (2016).

Kate Farrell and Judy Robertson. 2019. Interdisciplinary data education: Teaching primary and secondary learners how to be data citizens. In *Proceedings of the 14th WiPSCE Conference on Primary and Secondary Computing Education Research (WiPSCE '19)* (Glasgow, Scotland). ACM.

Birte Heinemann, Simone Opel, Lea Budde, Carsten Schulte, Daniel Frischemeier, Rolf Biehler, Susanne Podworny, and Thomas Wassong. 2018. Drafting a data science curriculum for secondary schools. In *Proceedings of the 18th Koli calling international conference on computing education research.*

William M Kapambwe. 2010. The implementation of school based continuous assessment (CA) in Zambia. *Educational Research and Reviews* 5, 3 (2010), 99–107.

Koby Mike, Tamir Hazan, and Orit Hazzan. 2020. Equalizing data science curriculum for computer science pupils. In *Proceedings of the 20th Koli Calling International Conference on Computing Education Research.*

Joshua A Muskin et al. 2017. Continuous assessment for improved teaching and learning: A critical review to inform policy and practice. (2017). `https://unesdoc.unesco.org/ark:/48223/pf0000255511`

Judy Robertson. 2019. Towards a sustainable solution for the shortage of computing teachers in Scotland. (2019). `https://www.research.ed.ac.uk/en/publications/towards-a-sustainable-solution-for-the-shortage-of-computing-teac`

Nathaniel Von der Embse, Justin Barterian, and Natasha Segool. 2013. Test anxiety interventions for children and adolescents: A systematic review of treatment studies from 2000–2010. *Psychology in the Schools* 50, 1 (2013), 57–71.