Raspberry Pi Foundation

Teaching primary learners how to be data citizens

Judy Robertson and Kate Farrell (University of Edinburgh)

Robertson, J. & Farrell, K. (2023). Teaching primary learners how to be data citizens. In Primary (K–5) computing education research – teaching and teachers. Understanding computing education (Vol. 4). Proceedings of the Raspberry Pi Foundation Research Seminars.

Available at: rpf.io/seminar-proceedings-vol-4-robertson-farrell

Teaching primary learners how to be data citizens

Judy Robertson and Kate Farrell (University of Edinburgh)

Introduction

Imagine the scene: the gym hall is full of excited ten-year-olds. They're rushing around comparing cards with pictures of coloured dragons printed on them[1]. Each card contains information about a single dragon — name, age, colour, size, and personality. Taken together, the set of cards is a sample of a population of dragons living in the wild. The children want to know whether red dragons are always angry. Are there any green dragons that breathe ice? Are the male dragons more likely to live longer? By organising these Dragonistics cards on the gym hall floor, they can create physical bar graphs to help them to investigate these questions. The teacher is delighted because the children are completely engaged in learning. She has been experimenting with adapting the activity to different age groups of learners and although she originally started with ten-year-olds, she is confident that six-year-olds "...can do it too. They just shout more."

If you're a teacher, you might be familiar with this sort of activity. Categorising and sorting information using physical objects and cards is commonly used to introduce computational thinking concepts about processing information[2]. It is also used in early years maths. However, the teacher's purpose in unleashing the dragons in the gym hall was to develop her learners' data literacy skills. Data literacy is about using data to answer questions and solve problems. It is often taught using an inquiry cycle called **PPDAC** (Wolff et al., 2016), which goes through stages of finding a **Problem** to solve, making a **Plan** for how to answer the question using data, collecting and organising **Data**, **Analysing** that data, and then drawing **Conclusions** (see Figure 1 for an illustration).

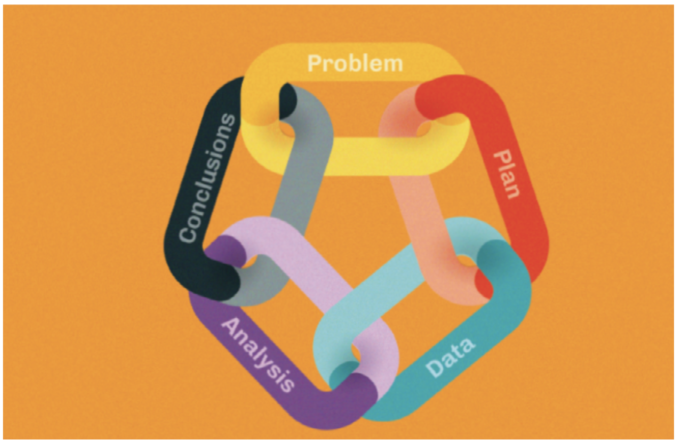

Figure 1. The PPDAC cycle.

There is an overlap between data literacy and other sorts of literacy, including statistical literacy and computational literacy (Bhargava, 2019), which is why it is useful to draw on familiar approaches from maths or computing. It is not just a technical discipline, though — it draws on similar skills to media literacy, for example, in sniffing out fake claims in the media by spotting where data is misused. Ethics is also an important aspect of data literacy; for example, considering how data can be used for societal good and paying attention to whether it has been collected in a fair and transparent way.

The purpose of teaching data literacy skills is to enable children and young people to be data citizens so that they can use data to make informed decisions in their daily lives and to make positive changes in their community. Part of data citizenship is about understanding personal data trails — what data is generated by the technology you use, including fitness trackers, online shopping sites, social media, and smart meters?

[1] You can order a set from https://creativemaths.net/our-creations/dragonistics/

[2] There are some unplugged activities in our TeachCS for Primary Practitioners Handbook at https://teachcs.scot/

A data-literate citizen might wonder: who collects that data, with whom do they share it, and for what purpose? Data literacy skills are the foundation for data science; some children might go on to become data scientists in future careers and address a pressing shortage in the labour market. But for us, that isn't the primary purpose of data literacy. The real purpose is to help children make sense of the complicated world in which they live today.

We don't mean to stress teachers by suggesting they should squeeze even more into the busy curriculum. We reviewed the Scottish Curriculum for Excellence and found that outcomes and expectations relating to the PPDAC cycle were already present across maths, English, social sciences, and technology-based curricular areas (Farrell & Robertson, 2019). There was an even wider scope to apply data literacy skills in projects in other curricular areas such as art or health and wellbeing. It may be that data literacy is spread across the curriculum in the country in which you teach. In our work with primary school teachers in Scotland, it has become clear that they often teach data literacy skills here and there across the curriculum, even if they don't call it by the same name. In one of our courses for primary teachers, we brought up an example of how the PPDAC cycle could be used to investigate birds in the school playground. "Oh, we do that!" a teacher told us, "But that's just the gardening club". Other examples quickly emerged of creating information displays of active travel data or sugar in soft drinks, or graphing physical activity during circuit training.

In this article, we describe the PPDAC data problem-solving cycle in more detail through a worked example of birdwatching in the school playground. We hope that it will help teachers to relate it to their own practice and perhaps experiment with new ideas. We chose this example to illustrate that data literacy is about intellectual skills and so it doesn't necessarily require technology. Technology is always useful though, particularly when the data sets get bigger, so the subsequent section introduces examples of data literacy projects that use digital tools.

Previous data literacy research

What should children learn about data? Previous researchers have quibbled about this question. Because data literacy is at the intersection of various subdisciplines, researchers from the subdisciplines sometimes grumpily point out that data literacy isn't new — it's just what they've been doing for the last twenty years but with a different name. Gould, a statistics educator, advocates for "data-scientific thinking", which "consists of a strong core of statistical thinking, carefully selected components of computational thinking, and just a dash of mathematical thinking" (Gould, 2021; p.12). For Gould, inquiry cycles such as PPDAC are "at the heart of statistical thinking" (Gould, 2021; p.14), and supporting learners to pose statistical questions as part of this process is particularly important. For example, in the gym hall example with the dragon cards, the teacher might help the learners to shape their initial curiosity into questions that will lead to a fruitful investigation (e.g., the question "Are male dragons in our card set likely to live longer?" could be the start of an interesting inquiry). See Arnold (2021) for an example of a classroom activity to develop statistical questioning.

Gould notes that another key aspect of data literacy is to be able to reason about the presence of chance variations in the data. To put it another way, learners need to be able to recognise that sometimes apparent trends in data can be caused by randomness rather than any underlying explanation. Think about an example where the learners noticed that the male dragons seemed to be older. Could it be that we just happened to pick lots of cards with elderly male dragons? If we looked at all the male dragons in the box, would the age pattern hold? What if we checked another box of cards? If we take more samples of dragons, and we notice the same pattern, then we can be more sure that the trend did not happen by chance.

Burrill and Pfannkuch (2023) reviewed the recent challenges and changes to the way statistics is taught. They identify that the availability of new software and data science techniques offers wider possibilities for learners to engage with statistics in a hands-on way. In particular, interactive visualisations enable learners to explore concepts that would have been previously hard to grasp, such as distribution and variation. They observe that "because technology can remove learning barriers (e.g., constructing graphs by hand and mathematical procedures), *all* students can engage in activities that are authentic to the discipline [of statistics], access statistical ideas from a very young age and experience reasoning from and about data and modelling" (Burrill & Pfannkuch, 2023; p.5).

Burrill and Pfannkuch also recognise the possibilities for new statistical learning contexts brought about by widespread and easily accessible data generated by technology. Learners need no longer painstakingly collect their own data every time, and nor do teachers need to make up fictional data sets. They can access large public data sets online or use sensors to generate data (see the DataFit and Internet of Things in Schools sections of this paper). Data can come in various formats: images, text, dates, and locations as well as traditional numerical values. The authors also underline the value of social statistics and how it can be used to address social justice issues. Statistics education increasingly needs to develop students' ability to interrogate data, know what questions to ask, reason about uncertainty, and advocate for social change.

The necessity of social statistics is also important to Weiland (2017), who proposes a framework for critical statistical literacy in which learners are encouraged to use their understanding of statistical concepts to analyse and critique social issues. He makes the distinction between being a critical *consumer* of data (e.g., being able to make sense of graphs in news stories) and being a *producer* of data to use an inquiry cycle to examine meaningful real-world data and use it to make changes. As an example of a project where learners were producers of data in this sense, a teacher recently told us about a project in which the children used the PPDAC cycle to explore school lunch packaging, persuading the local authority to reduce the amount of plastic they use.

Weiland's view of critical statistical literacy has some common ground with Pangrazio and Selwyn's critical data education (Pangrazio & Selwyn, 2021). However, they focus in particular on the issue of the personal data trails that are generated when we use apps and gadgets. There have been concerns about the datafication of society, which refers to the vast amount of data that is collected on digital devices, the automated analysis performed on this data, and the consequences for societal inequalities, which can be perpetuated by decisions that are made on this basis.

Rather than focusing on inquiry or questioning with data, Pangrazio and Selywn educate learners about the ways in which technology companies use personal data and the implications for this in our technological society. During critical data education, young people learn about how data about individuals is collected and processed, critically examine how metrics are used in social media use, and learn strategies for protecting their personal data. The purpose is to get learners to "actively consider and question digital content and proactively manage the implications of their practices" (Pangrazio & Selwyn, 2021; p.444). Although awareness of how personal data is used (and abused) by technology companies is an important topic, our preference has been to use the PPDAC data inquiry cycle to develop learners' skills in being consumers and producers of data.

Working through the data problem-solving cycle unplugged

Our team — Data Education in Schools — developed a set of data explorer cards as hands-on, non-digital activities on four different themes. Each theme has five cards, one for each stage of the PPDAC cycle[3]. We intend for teachers to give groups of children the cards to prompt their thinking. Every group could work on the same card, or they could choose different stages of the cycle. We'll look at the bird-themed project in this article. The prompts on each card could easily be adapted to different themes if you know your learners are interested in a different topic.

[3] Classroom resources to support this example are available here: https://dataschools.education/resource/dataexplorercards/, along with other examples including Lego, plastics, and lost property.

Problem

This card helps the learners to narrow down what they would like to investigate relating to the theme of birds in their school playground (see Figure 2). The learners are asked what they notice about birds in the playground and what they might want to investigate more. The prompts are intentionally open-ended to provide a loose structure to guide thinking without constraining it too much. By the end of the Problem stage, the learners should have a question that they would like to answer about the birds that visit their playground.

Figure 2. Example task for the Problem stage of the Birds project.

Plan

The purpose of the Plan stage is to work out how best to answer the question that the learners identified in the previous stage. The prompts encourage the children to consider the possibilities for the sort of data they could collect to answer their questions about birds and to look ahead to the sorts of answers they expect or the challenges they might encounter (Figure 3). Here, the teacher has the balancing act of encouraging the learners to be independent in their planning, while making sure that their plans are feasible and will enable them to answer their questions.

Figure 3. Example task for the Plan stage of the Birds project.

Data

At the Data stage, the learners collect, record, and organise their data. The photos on the card give some hints about techniques that the learners could use — binoculars for observing birds, tally marks in a jotter for recording different bird species, a map that shows the locations of different bird feeders, or a bird feeder that is annotated with the amount of bird seed left on each day (see Figure 4). There are opportunities here for the teacher to link in with other curricular areas, e.g., maths for counting with tally marks or measuring quantities of seed.

Figure 4. Example task for the Data stage of the Birds project.

Analysis

The prompts on the Analysis card encourage the children to look at their data and *notice* and *wonder* what is going on (Figure 5). Again, the photos give some clues about the sorts of things that they may notice in their own data (if they collected data from bird feeders). The key here is to support the children in making sense of patterns or trends in the data using simple representations. Later in their secondary school or college careers, they might learn fancy statistical analysis techniques, but they don't need those now. What they need to develop is a secure understanding of critical reasoning about data presented in physical objects, tables, and graphs. They need to be able to read these representations to help them find the answers to their questions and to compare possible explanations that the data could give them.

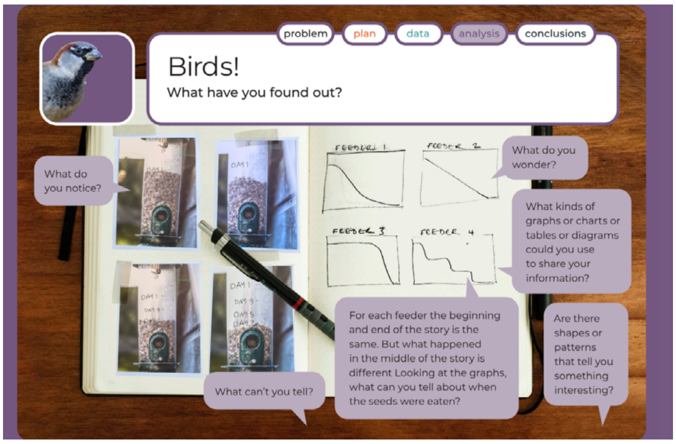

Figure 5. Example task for the Analysis stage of the Birds project.

Conclusions

This card brings the learners full circle in the PPDAC cycle and prompts them to answer the question they posed in the Plan stage. The prompts guide the learners to consider what they found, whether their findings are what they expected, and if there are limitations to the method they used (Figure 6). They may be inspired to answer a new set of questions and want to start a new cycle. Perhaps the findings will encourage them to make changes in the school — maybe the bird feeder needs to be positioned somewhere that the janitor's cat can't reach, or maybe they decide that planting more wildflower seeds will result in more feathered visitors next year.

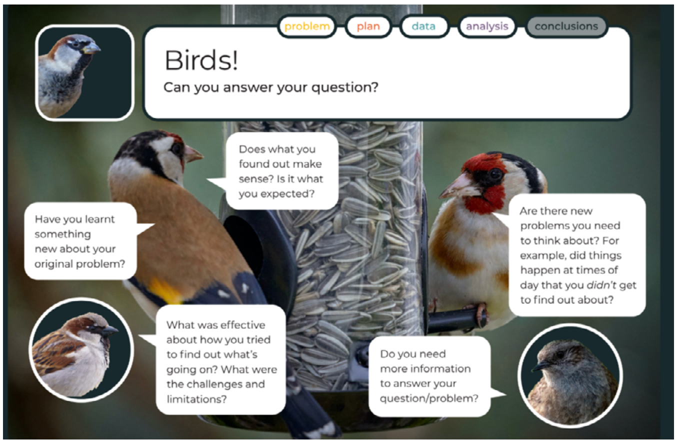

Figure 6. Example task for the Conclusions stage of the Birds project.

Using digital tools to support data literacy

We have seen that unplugged approaches — dragon cards or bird watching — can be a great way to build the foundations of data literacy, particularly in the early years. Digital tools can also be used to make it easier to deal with large quantities of data or to teach children new digital skills.

Data skills live lessons

We collaborate with a company called Digital Skills Education to produce interactive online lessons about data topics. The purpose of the live lessons is to motivate learners with an exciting example of how data can be used in interesting ways in the real world. It can be a springboard to further data lessons in the class. Teachers can sign up for free for a live event in which their learners can participate, or they can choose to follow along with the activity using a recorded video if they prefer.

One of the most popular lessons is called Defend the Rhino with Data Science. This lesson has a storyline in which a park ranger from a fictional wildlife park asks for the participants' help in locating where rhino poachers are likely to strike next based on photos from the park's CCTV cameras (see Figure 7). Because there are so many cameras, which generate so many photos, it becomes clear that even if the participants work as quickly as they can, it is simply not possible for humans to do this.

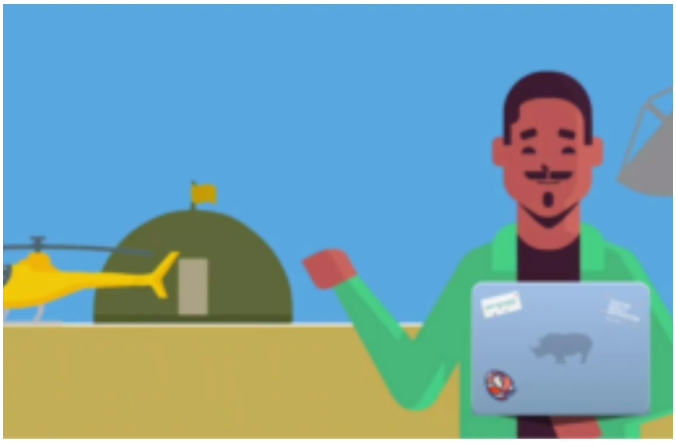

Figure 7. A screenshot of the park ranger from Defend the Rhino with Data Science.

The presenters then introduce the idea of machine learning — in this case, training a computer to automatically classify what is in an image based on photo data. The learners are asked to help train a machine learning model by labelling images that contain humans and those that contain only animals. In the story, the park ranger sets the machine learning model to work, and the learners use the output from it to figure out where the poachers are most likely to be. He then sends a helicopter to stop the poachers, just in time!

During the lessons, the learners watch the video of two presenters explaining a scenario, and as learners complete online tasks, the presenters give a commentary about how different schools are doing ("Well done, I see that Tinto Primary School has worked out where the poachers are"). In other lessons, participants learn how to do simple coding to generate images that represent information about themselves (Data Selfies) or calculate the area of their school roof and use NASA satellite data in order to investigate whether installing solar panels would be financially viable (Plug in the Numbers).

These lessons are popular — we have had 29,000 learners from 27 countries take part in the last three years. You can see an example video of Data Selfies here: rpf.io/data-selfies-video. Keep an eye out for upcoming live lesson events at https://dataschools.education/events/.

DataFit

Children often encounter data during interactions with everyday technology, so it is important to ensure that they know how to make sense of it and that they understand the privacy implications of what technology companies do with their data. Our DataFit project[4] is designed to teach upper-primary (aged 7–11) and lower-secondary (aged 11–14) school learners how to interpret physical activity information (such as step counts) from devices such as Fitbits or their phones (see Figure 8 for an example from a Moki fitness tracker). The visual displays used by fitness devices contain complex and often interactive data in a format that is not traditionally taught in schools. Teaching children how to read such displays (for example, with comprehension questions as in Figure 8) is useful to ensure that they can correctly interpret their personal data so they can use it to make informed choices. It encourages them to reflect on their health behaviour and consider how using fitness data can help them to change their habits. We developed this resource in conjunction with our colleagues who research physical activity for children's health, so it contains useful background information about why physical activity is so important. It is also beneficial to discuss with learners who has access to the data that their fitness device collects and how to change the data sharing settings to suit their preferences. This is an example of teaching children about their personal data trails (Pangrazio & Selwyn, 2021).

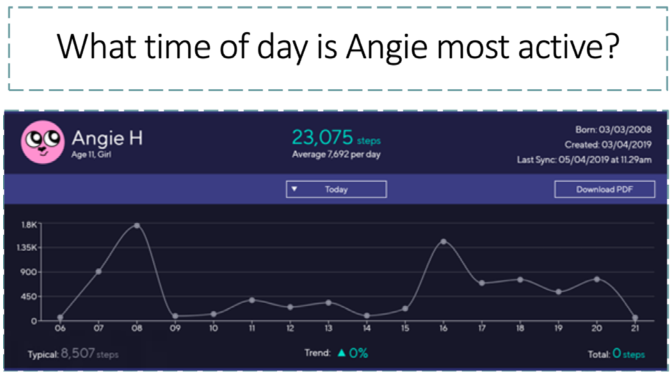

Figure 8. An example from a DataFit lesson that teaches learners how to analyse time series data from a fitness device.

Common Online Data Analysis Platform (CODAP)

At primary school, data analysis skills can be introduced through visual representations, even in the early years. The teacher's aim is to support learners to be able to switch between different representations fluidly and to help them use these representations to draw reasonable inferences. CODAP[5] is a free online tool that makes it very easy to switch between data shown in tables, dot plots, and other sorts of graphs. Learners can drag and drop different variables to quickly make graphs. Although spreadsheet software also does this (and some spreadsheet tools are free), CODAP has the advantage that it was designed for children and focuses only on data analysis. This means the user interface is considerably easier to learn than spreadsheet software aimed at adults.

There are many examples of real-world data sets and exercises on the CODAP website on science and social science topics (as recommended by Weiland (2017)). It is also possible to import or type in your own data. For example, you might want to use software to investigate further the properties of the herds of dragons in the Dragonistics card sets. The learners might start by representing the dragon data by sorting out the cards on the floor (see Figure 9, top). But if they wanted to examine data from a larger sample of dragons, this would take a long time. This is an ideal opportunity to demonstrate how software is useful for handling larger data sets quickly while also bridging between physical representations and more abstract representations. The bottom image of Figure 9 shows a CODAP graph that we made based on a Dragonistics data file, which we downloaded from the Creative Maths website[6]. Note that the dot plot style enables learners to make the connection between dots and single data points in a similar way to the physical cards. CODAP also enables the dots to be merged into the more commonly seen bar chart once learners have grasped the connection.

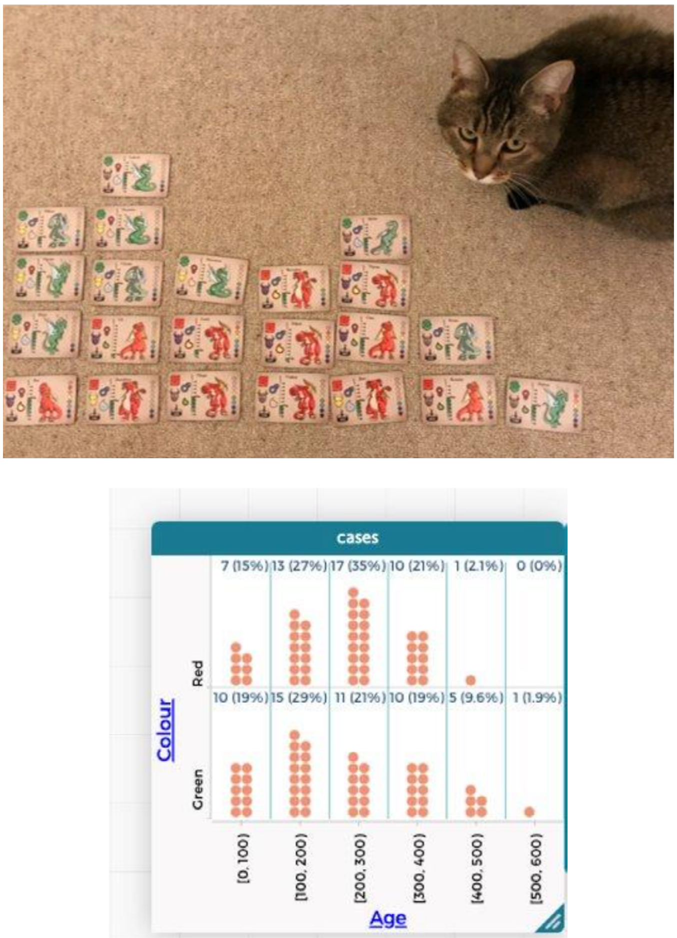

Figure 9. Dragons cards arranged by colour and age physically (top) and using CODAP (bottom).

Internet of Things in Schools

If you are feeling adventurous about technology, you could consider using environment sensors installed in the classroom to teach the children about data. This is the approach we have taken in our Internet of Things in Schools project[7]. The University of Edinburgh has installed Internet of Things sensors in 34 primary and secondary schools, which measure CO_2, temperature, humidity, air pressure, light levels, and movement. The sensors send data over a private network to secure university servers and the learners can see the data graphed in real time on a webpage. In the hands of some enthusiastic teachers and learners in our pilot schools, these sensors have led to some great examples of curiosity-led learning (Robertson et al., 2023). Learners in Roslin Primary School noticed the impact of the Tongan earthquake on their air pressure sensor and compared their time/speed/distance calculations about how fast the pressure wave was moving with a university scientist who was studying the eruption too. At Addiewell Primary School, the learners persuaded a local zoo to install a temperature and movement sensor in the crocodile enclosure because they were curious about the environments that different animals need. Lynburn Primary learners investigated how CO_2 levels impact how ready they feel to learn, and whether CO_2 levels can be lowered by introducing plants into the classroom. In these examples, the learners worked through the stages of the PPDAC cycle, but in the Data stage, the data collection and visualisation were primarily done automatically by the sensors. This is a powerful illustration of the strengths of machines and humans during PPDAC — sensors and computers are useful for regularly collecting accurate objective data from the environment; the role of humans is to interpret that data and use it to solve problems.

Our Internet of Things in Schools project has funding to work with schools in South East Scotland, but don't despair if you work elsewhere: the SAMHE project[8] is currently offering free air quality monitors for UK schools and these sensors have a similar purpose to the sensors we used in our project.

Conclusion

Let's go back to the gym hall to revisit the children with their dragon cards. We now know that the chaos was actually the PPDAC cycle in action. Having played with the cards before, the children had noticed that the red dragons seemed to be angry more often than the greens. They wanted to find out whether this was true or not (PPDAC – Problem). The teacher helped them to plan how to investigate this by investigating the dragon population across a whole card pack (PPDAC – Plan). Sorting the cards physically enabled the children to organise the data (PPDAC – Data) while arranging the cards into a bar graph was analysis (PPDAC – Analysis). The teacher might, in the future, choose to show the learners how to use a digital tool such as CODAP to investigate larger sets of dragons. In discussing the graph with the other learners and their teacher, the children drew conclusions to answer the question (PPDAC – Conclusions). Curious readers might be satisfied to learn that red dragons are indeed more likely to be angry than their green counterparts but that angry green dragons do exist. Data-literate readers will want to check these claims for themselves.

One teacher who taught with the dragon cards vowed to "never teach graphs with a textbook again". We hope the resources we mention in this article will tempt you to leave your textbooks to gather dust too.

Resources

If you would like to find out more about teaching data literacy, download our handbook from https://dataschools.education/teachdata.

References

Arnold, P. (2021). Questioning through the Statistical Problem-solving Process. *Statistics Teacher/STatistics Education Web: Online Journal of K–12 Statistics Lesson Plans*. https://www.statisticsteacher.org/ or http://www.amstat.org/education/stew/

Bhargava, R. (2019). Data Literacy. In *The International Encyclopedia of Media Literacy* (pp. 1–5). Wiley. https://doi.org/10.1002/9781118978238.ieml0049

Burrill, G., & Pfannkuch, M. (2023). Emerging trends in statistics education. *ZDM – Mathematics Education.* https://doi.org/10.1007/s11858-023-01501-7

Farrell, K., & Robertson, J. (2019). Interdisciplinary Data Education: Teaching Primary and Secondary learners how to be data citizens. In *Proceedings of the 14th Workshop in Primary and Secondary Computing Education* (pp. 1–2). ACM. https://doi.org/10.1145/3361721.3362120

Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, *43*(S1), 11–22. https://doi.org/10.1111/test.12267

Pangrazio, L., & Selwyn, N. (2021). Towards a school-based 'critical data education'. *Pedagogy, Culture and Society*, *29*(3), 431–448. https://doi.org/10.1080/14681366.2020.1747527

Robertson, J., Amirkhanashvili, G., Abaci, S., Linklater, H., & Lawson, T. (2023, June 21). *Learning about data literacy pedagogical practices from primary school teachers*. EdArXiv. https://doi.org/10.35542/osf.io/wsn8p

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics, 96*, 33–47. https://link.springer.com/article/10.1007/s10649-017-9764-5

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, *12*(3), 9–26. https://doi.org/10.15353/joci.v12i3.3275

Judy Robertson
(University of Edinburgh)

Professor Judy Robertson is a professor of digital learning at the University of Edinburgh. She is the academic lead of the Data Education in Schools project, which aims to promote data literacy in school learners in South East Scotland. She has been developing educational technology with children and teachers since 1997. Her work focuses on how technology can help to solve thorny real-world problems, including serious games and apps for education and health.

Kate Farrell
(University of Edinburgh)

Kate Farrell is an educator with expertise in data science, computing science, digital media computing, and moving image education. She advised on the new national computing science curriculum in Scotland and wrote guides for teaching computing science for early years, primary, and secondary practitioners. She currently works on the Data Education in Schools project, developing data science and data literacy lessons, resources, and professional learning for primary and secondary schools in Scotland.

Raspberry Pi
Foundation

www.raspberrypi.org

Raspberry Pi Foundation, UK registered charity 1129409