Data education and mathematics 5-16: a rapid review for the Joint Mathematical Council of the UK and the Royal Society

Vinay Kathotia and Cathy Smith
Open University 2026

The Open University

Contents

The Open University

1. Introduction

This report follows a rapid review of key source documents that have informed and communicated the work of the Joint Mathematical Council of the United Kingdom (JMC) and the Royal Society (RS) with regard to data education. The rapidity has been necessitated by the desire to inform the current revision (DfE, 2025b) of the National Curriculum in England, for students aged 5 to 16, which is running to a compressed schedule (drafting of mathematics Programme of Study in March 2026 for a revised national curriculum in 2027 and first teaching in 2028).

One consequence of this constraint has been that we have consulted a small corpus of evidence in drawing up the curricular competencies for data education that are the main content of this report. We briefly discuss this and other limitations below, in the section on the **Method for this review**.

Despite this substantial constraint, the rapid review provides an opportunity to impact and improve the provision for data education for students in England. It has two purposes:

- To add specificity and granularity to the visions for mathematical and data education that are set out in recent JMC and RS reports (notably, Smith *et al.*, 2023; Jacques and Joubert, 2024; The Royal Society, 2024) in the form of statements that can be compared with existing and future curricula;
- To support those drafting and exemplifying the curriculum in England, by proposing statements that enhance progression and coherence.

While curricular statements in a Programme of Study represent granular content and processes of disciplines, they do not fully capture their detail nor the wider, holistic aspects, including how the curriculum is interpreted, supported, enacted and experienced. To clarify some of these ideas, we have included a section on **Ideas underpinning data education within/alongside mathematics**. We have also included a section on **Implications for the current revision of the National Curriculum in England**, which is informed by principles (e.g. connectedness, sequencing, specificity), set out in the government's response (DfE, 2025b) to the Frances Review (DfE, 2025a), which guide the current revision of the National Curriculum in England.

Acknowledgements

We would like to thank Jane Clarke, Catherine Boulton and Noel-Ann Bradshaw, for outlining the scope of this rapid review and making it possible, and the authors of the various reports, research articles and curriculum documents whose work we have set out to synthesise here.

2. Method for this review

This review builds most directly on two reports: Smith et al.'s (2023) report for The Royal Society (RS) which sets out a framework of competencies and toolkits in mathematical and data literacy for all/some/many school leavers, and Jacques and Joubert's (2024) report (hereafter referred to as JJ) for JMC that adapted that framework and used an analysis of the four nations curricula to add benchmarks at the end of primary, end of lower secondary, and at school leaving age.

Following the Curriculum and Assessment review, England's government set out a desire for:

- a well-sequenced Programme of Study (PoS),
- which clarifies connections within and across subjects (to be made available via an online "3-D" interface), and
- has greater specificity than the current PoS.

We have thus:

1. Taken the structure of competencies and toolkits proposed by Smith et al. (2023) and then refined by JJ.
2. Compared these with more detailed progressions from a specific set of curricular sources (see list below) identified in those reports and by the RS and JMC as products of extended design work involving STEM professionals, HE educators and teachers
3. Cross-referenced these to the mathematics PoS in England's existing NC (DfE, 2013a; DfE, 2013b; DfE, 2014), and

4. Identified a more granular progression that more carefully details the required data skills. The granularity has been determined by wanting a granularity as close as feasible to England's existing NC documents (currently year by year at Key Stages 1 & 2, and then by Key Stage) but being mindful of the granularity of our source progressions (which can be as coarse as 2 or 3 years). These sources range from specifying three or four competencies per school year to over 150. The rationale is to retain a level of detail necessary to raise awareness of the competencies' application to statistical thinking, and to draw out related progressions (see accompanying spreadsheet). Where possible, we have indicated examples of contexts and activities suitable for the relevant age range and, at the bottom of each row, we note some of the rationale for what is included.
5. Considered how a coherent and holistic approach to data education can be fostered alongside specific competencies. This informs the following section on Overall Principles.
6. Furthermore, England's current NC Programmes of Study (in mathematics and in data-related subjects such as Computing, Geography, PSHE and Science) have data-related competencies in the subject content, in some non-statutory guidance (NSG), and in preambles preceding the listing of content by Key Stage, all of which can add important detail. We have therefore also touched on some of these elements in the subsections/tables of our **Curriculum guidance** section.

Limitations

The 'evidence' we draw on in this report is not primary research evidence. We are aware of no trials that demonstrate the superiority of one approach to curriculum inclusion and organisation over another, and certainly not in the new field of data education. Given the rapidity of the review, we have not sought results of finer-grained experiments.

Instead, we have drawn on sources that have been implemented at scale, and for some years, e.g. the curricula of New Zealand, Scotland and Singapore. While none of these approach the scale of England's education system, they do operate across a range of schools. There are problems in comparing performance internationally (not least that this ignores known effects of social disadvantage) but we note that these sources cover a range. In PISA 2022, students in Singapore performed better at uncertainty and data questions than English students do, and English students outperformed New Zealand and Scottish students (Ingram et al., 2023). This is welcome but, in England, out of the four PISA mathematics domains, uncertainty and data showed the largest gap between the highest- and lowest-performing students, indicating that many students are being left behind in data education.

As well as scale, the other quality of the sources we use is their roots in iterative design research involving teachers, researchers and statistics curriculum developers. GAISE II (Bargagliotti et al., 2020), the New Zealand (NZ) curriculum (Tahurangi, 2026), Scotland's Data Skills and Data Literacy curriculum (Data Education in Schools, 2024) and the Data Education for Everyone (DE4E) data science learning progressions (Data Science Learning Progressions, 2025) are notable examples.

A related limitation, that we have not addressed, is that we have not considered the connections between mathematics and AI that are less relevant to statistics. Notably, AI and Machine Learning (ML) make use of mathematical concepts such as vectors, circles and decision trees, and there is design research (e.g. ProDaBi, please see Appendix 2) trialling lessons for school-age children that make these connections explicit.

Key Sources	
Abbreviated ref	**Reports**
(RS)	• *Curriculum and Assessment Review* (DfE, 2025a) • *Government response to the Curriculum and Assessment Review* (DfE, 2025b) • *Mathematical and Data Literacy Competencies and curriculum implications at the intersection of mathematics, data science, statistics and computing* (Smith et al., 2023), for the Mathematical Futures Programme • *Educational Technologies in mathematics education* (Crisan et al., 2023) for the Mathematical Futures Programme • *The Integration of data science into the primary and secondary curriculum* (Pittard, 2018) for RS ACME

Key Sources	
(JJ) (PEY) (RSS)	• *A vision for meaningful data education across the four nations of the UK*, JMC (Jaques & Joubert, 2024) • *Data Education – Progression in the primary years* (ACME PEY Expert Panel, 2025). • RSS report: *Key Recommendations for the UK Statistics Curriculum.* (Jones & Dudzic, 2024).
Abbreviated ref	**Curriculum documents**
(NC) (NZ) (DE4E) (GAISE II)	• England • New Zealand • Scotland • Singapore • *Data Education For Everyone* (DE4E) project learning progressions (Data Science Learning Progressions, 2025) • GAISE II report from the US (Bargagliotti et al., 2020)

3. Overarching principles for data education within/alongside mathematics

The scope of data education

One issue that we face is matching the concept of data education with the structure of the school curriculum, and with our particular interest in mathematics.

Data education can – and should - take place in many school subjects and could transcend school subjects. Within the four nations of the UK, Wales provides an example of a curriculum (https://hwb.gov.wales/curriculum-for-wales) where subject knowledge is developed in service of the curricular aims of informed citizenship. Scotland's curriculum is developed in subjects but data skills have been extensively mapped and cross-referenced across subjects. England's curriculum is subject-rich; solving problems with data is mentioned in science, geography and, implicitly, in mathematics. The stance we take is that, wherever data education is taught, the mathematical and statistical elements of it should be made explicit and that this will not happen unless they are specified in all the relevant curriculum guidance and have coherence with the underpinning concepts in mathematics.

The model of data education (DE) developed in Smith et al. (2023) (Fig 1) positions data education as an overarching activity of critical enquiry, applicable in any domain or school subject, that is centred on posing and solving data-driven problems and is supported by mathematical, computing and statistical/data toolkits. (Note that it argued, following many statistic educators, that statistics education, were it to embrace the growing role of data and technology, it would be equivalent to data education at school level.)

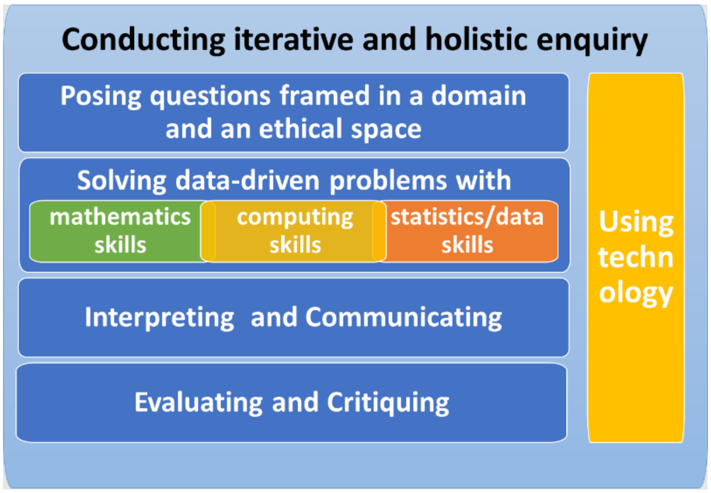

Figure 1 Model for data education (which could also encompass mathematics)

Jacques and Joubert (2024) adopt the term 'data skills' to describe the competencies learners need to achieve 'data literacy.' Although similarly focused on mathematics, they use a compatible definition given by Wolff et al. (2016), which is implicit about mathematical and computing skills, and emphasises storytelling and design:

> *the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyse, visualise, critique and interpret data, as well as to communicate stories from data and to use data as part of a design process. (p. 23)*

The broader vision of data education proposed in these documents, with its contextual relevance and ethical aspects, provides the opportunity for an engaging, data-driven education that includes all the statistics needed by school leavers.

Where could data education sit?

Statistics education historically sits within mathematics education. There are well-rehearsed disadvantages to this approach. Statistics educators and employers argue that posing, interpreting and solving problems with data in a relevant context, for a purpose and with appropriate technological tools, are central to statistics/data science. Notably, context is not something that can be abstracted away at an early stage but informs the full cycle of statistical enquiry (Smith et al., 2023). Further, Jacques and Joubert's review of the UK national curricula concludes that current UK expectations in mathematics/statistics are too narrow for workplace needs and that they "ignore the broader skills associated with data literacy, critical thinking and awareness of ethical data use (data stewardship)" p22. These skills could usefully, and probably more easily, be integrated into citizenship, science, humanities and oracy.

Data education also requires a fluent use of technology that is clearly related to computing (which includes programming). There is not currently sufficient capacity in school computing to include the RS/JMC vision for data education. We consider, however that many of the principles and curriculum decisions that we discuss, in the light of mathematics, would also apply for computing and that similarly detailed guidance would be necessary for computing teachers to teach its statistical elements.

The work of The Royal Society and JMC argues for a recognition of data education throughout the curriculum and not its removal from mathematics. Instead, documents like this aim to support governments in refining the statistical content across the curriculum to show progression in the skills and experiences that build data literacy. It is worth repeating that, wherever data education is taught, the mathematical and statistical elements of it should be made explicit and that this will not happen unless they are specified in all the relevant curriculum guidance and have coherence with the underpinning concepts in mathematics.

Principles of data education

These core principles are abstracted from the source documents. They serve to bring coherence to the granular statements that follow, and to justify them.

1. Students should meet real data, i.e. data that comes from a relevant context, is of a meaningful size and has initial complexity. (RSS)
2. Students need to engage in the full investigative cycle, including posing and refining questions, leading to obtaining, evaluating, refining and communicating solutions, and involving iteration as necessary. (all sources)
3. Working with data is not just statistical analysis. It includes data stewardship, handling data, applying statistical tools/analysis/thinking, and representing, visualising and communicating data. (all sources)

4. Statistical thinking involves dealing with variability – dealing with chance. Statistical data representations serve to represent and help interpret variability. (all sources, Kathotia, Ward-Penny & Smith, under review)
5. Uncertainty is a foundational principle. Mathematical/statistical methods enable us to accept uncertainty, treating it with increasing precision and confidence. (RS; Kathotia, Ward-Penny & Smith, under review)
6. Use of technology is integral to data education. Data education addresses complexity using technology *and* using the power/simplicity of mathematics. (all sources)

In addition to underpinning our framing and choices for the curriculum guidance that follows, these overarching principles could also be useful for curriculum drafters, for the exemplification work to be done by organisations such as the Oak National academy, for teacher professional development, and for a reappraisal of our mathematical beliefs and values.

4. Curriculum guidance

In this section we set out data education competencies for Early Years, Years 1-2, 3-4, 5-6, 7-9 and 10-11. We follow the framework of earlier RS/JMC reports (RS and JJ) in describing the data education toolkit as a combination of:

- Statistical (holistic and iterative) Enquiry
- Data Stewardship
- Handling Data
- Statistical thinking and methods
- Probabilistic Reasoning
- Data representation
- Using aids and tools

Within each age range, and each competency we provide detailed statements of what students should be taught that, taken together, characterize that competency. When read across the different pages, they describe progression through the curriculum in each competency.

The sequencing of these statements, and in some cases their wording, is derived from our analysis of the source documents. We claim no originality in the wording of these statements. They were synthesised from the NC for England and the key sources listed in the methodology section. In cases where the statement draws most directly from only one source, this is indicated by the relevant initials to allow reference back to those sources. Each competency row also includes a 'Why' statement that briefly explains the decisions made.

Given that one role of this document is to inform England's curriculum review, we have started each section with the relevant part of England's current mathematics curriculum and aligned our wording with this where possible, for example by using the preface 'Pupils should be taught to …'. We have also included what we consider relevant parts of the citizenship (PSHE/RSHE), computing, geography and science curricula. Our competency approach to mapping data education inevitably results in a longer and more detailed description than is currently offered in England's curriculum.

We emphasise that these competencies should not be taught as a discrete list but as complementary aspects of statistical thinking and problem solving. Many are implicit in England's existing programme of study and are already included in good statistics teaching. We have also indicated in a lighter blue font where the statements indicate knowledge that is likely to be new to England's curriculum for mathematics. The purpose of specifying the competencies separately is to support all teachers in giving students experiences of these aspects of data education. They provide a means of checking that a lesson or activity has sufficient breadth and depth to allow students to progress in data skills and not to repeat superficial experiences.

Early Years

We do not make any specific recommendations for **data skills** in the early years, but we do advocate **purposeful** and **relevant** experiences of working with data throughout the early years just as for other phases of education. (JJ)

In the early years, children's enquiries which require data collection are rare and it would be unhelpful to be answering contrived questions. (adapted from PEY) That said, they can meaningfully experience elements of the PPDAC cycle.

Young learners are naturally curious about their environment. They gather information and make sense of it by playing with, exploring and sorting collections of objects and or repeating events to establish patterns or solve problems. These experiences are the beginnings of statistical inference (e.g. Makar and Rubin, 2020). As for all learners, data should be situated in the learners' natural environment. Synthetic experiences such as gathering information about eye colour etc. which do not begin with **purpose** (i.e. to assist in solving a problem or making a decision), will be a meaningless exercise and potentially lay the foundation for negative attitudes towards working with data. Stories can form a meaningful context for gathering data e.g., organising and representing the number of animals that join the old lady in *A Squash and a Squeeze*. (JJ)

The above approach ties in well with the area of 'Understanding the world' in England's EYFS statutory framework, which includes engaging with similarities and differences, drawing on their experiences, readings and storytelling.

Primary

In the primary years, the main focus of data education should be the full data enquiry cycle e.g. question, plan, data collection, representation, analysis and conclusion (not just kinds of representation). This needs to be in contexts where there is genuine enquiry (i.e. outcomes matter to children) e.g. voting for an activity, improving their own performance, testing their prediction in an investigation. Children need to experience data as meaningful, purposeful, and connected to real questions about their world. While the data cycle is constant, data thinking progresses in complexity with age. (PEY)

Additionally, the early and primary years investigations and embodied activities can serve as 'anchoring experiences' (PEY) for particular learning progressions and concept development. For example, experiences with sorting objects in a variety of ways can seed notions of Venn diagrams and two-way tables. Please see Appendix 1 for some of the anchoring experiences and concept progressions.

Note: the full data enquiry cycle mentioned above aligns with the PPDAC cycle: Problem, Plan, Data, Analysis, Conclusion.

DE4E and other sources see *Storytelling with data* as an essential component of the statistical enquiry cycle *at all stages*. They have a fair few related progressions (developing narratives, choosing appropriate visualisations, adapting for audiences, …) but we have included a limited thread here.

During Years 1 and 2

England's NC statement	Starting in Year 2, pupils should be taught to: • interpret and construct simple pictograms, tally charts, block diagrams and simple tables • ask and answer simple questions by counting the number of objects in each category and sorting the categories by quantity • ask and answer questions about totalling and comparing categorical data. Non statutory: Pupils record, interpret, collate, organise and compare information (for example, using many-to-one correspondence in pictograms with simple ratios 2, 5, 10).
Statistical enquiry Y1-2 (and sub-strand on Posing questions)	[Link with Science] The Years 1-2 strand on 'Working scientifically' in England's Science NC includes 'asking simple questions and recognising that they can be answered in different ways' and 'gathering and recording data to help in answering questions.' The (non-statutory) Notes and guidance state 'Pupils in years 1 and 2 should explore the world around them and raise their own questions. They should experience different types of scientific enquiries, including practical activities' [Use] the statistical enquiry cycle to understand issues relevant to their immediate locality and context, using primary data they have collected. (JJ) Year 1: Use data to ask and answer investigative questions with limited categories, e.g. do students in our class have one foot longer than the other? (adapted from NZ) Year 2: Use categorical data for an investigative question with limited categories (e.g. What are the favourite pets of students in our class?)(adapted from NZ) Posing questions: Ask Questions That Data Can Answer Years 1-2: Have students vote on a class question like "What's our favorite playground activity?" Then ask: "How could we find out if other classes feel the same way?" Help them think about who they'd ask and how. (DE4E) Why? The statistics education literature (e.g. GAISE II) underscore the importance of situating statistical learning in meaningful statistical investigations.
Data Stewardship Y1-2	[Link with PSHE/RSHE] If/when gathering data from fellow students, understand that collecting [and sharing] data about people requires their permission. e.g., asking before writing down a classmates favorite colour' (Ref DE4E) [Link from England's Computing PoS, KS1 subject content] Pupils should be taught to use technology safely and respectfully, keeping personal information

	private; identify where to go for help and support when they have concerns about content or contact on the internet or other online technologies. [Link from England's PSHE PoS, KS1 Media literacy and Digital resilience section] Pupils learn L7. about how the internet and digital devices can be used safely to find things out and to communicate with others L9. that not all information seen online is true Why? None of the surveyed mathematics curricula had mention of data stewardship or touched on sensitivity, privacy, reliability, data security, impact and responsibilities Have included points from DE4E progressions relating to ethics and risks which tie in with PSHE PoS and could be relevant for classroom investigations. Evidence base unclear.
Handling Data Y1-2	Year 1: [Link to measurement] Order objects by size or length. [Link to Measurement strand in England's mathematics Year 1 PoS] Pupils move from using and comparing different types of quantities and measures using non-standard units, including discrete (for example, counting) and continuous (for example, liquid) measurement, to using manageable common standard units. Collect, sort, count and record categorical data (e.g. colour of hair or toys) that is relevant to their immediate locality and context, with limited, but more than two, categories, e.g. means of travel to school or voting for class activity. Record data using tally charts. Simultaneously sort objects using two attributes, e.g. lay out a 2D grid on the classroom floor and divide toys by colour along rows, and type (animal, car, brick, …) along columns. Using a 2x2 grid to sort toys into "blue or not blue" (along rows) and "car or not car" (along columns) can seed the progression to two-way tables Additional in Year 2: Students record, interpret, collate, organise and compare data arising in practical activities and investigations. (adapted from England's non-statutory guidance) When sorting categorical data into categories, consider if 'other' should be a category for sorting rare responses (adapted from NZ) Why? Statements above are borrowed or adapted from England's and NZ curriculum. The processes/practices themselves are already in England's mathematics curriculum although stated for number, measurement or visual representations (such as pictograms). Given the material and embodied learning at KS1, would make more sense to have a data framing for activities

	such as 'collect, sort, order, count, organise, compare, interpret' providing tangible starting points for data learning progressions.
Probabilistic Reasoning Y1-2	[Link with Oracy] Make predictions with evidence and caution, using language of uncertainty (likely, unlikely, may be, because). Why? All sources defer formal probability since fundamental underpinning ideas of frequencies, equality and sharing need to be established first in number and spatial reasoning (PEY). Scotland, DE4E, GAISE11 emphasise early attention to expressing appropriate uncertainty when reasoning.
Statistical thinking and methods Y1-2	Collect and record multiple measures of the same numerical attribute using non-standard units and describe the variability e.g. use hands to measure tree girth. [Link to measurement] Ask and answer simple questions about frequency by counting the number of objects in each category. Ask and answer questions about totalling and comparing categorical data, using variable names and language of most/least/equal to/more than /less than/total [Link with number] Why? All sources work only with attributes in early years, building foundations in counting, comparing, ordering and adding numbers and communicating the meaning of the number. Variability is central to statistics, and can be connected to use of non-standard measures in Years 1 and 2, laying the foundations for further work. Language of frequency is important
Data representation Y1-2	Year 1: Use physical blocks to represent frequency for categorical data. (NZ) Create picture graphs for categorical data i.e. represent all available values for a variable and their frequencies (NZ) Use consistent and equally-sized images (or objects) to represent frequency (adapted from NZ) Year 2: Interpret and construct simple pictograms, tally charts, block diagrams and simple tables (England's NC) Use and interpret dot plots which represent each data point with a dot of the same size. (adapted from NZ) Describe data represented in very basic ways and begin to link observations. (DE4E)

	Why? NZ starts progression from pictograms to dot plots in Year 2. (Given relation with block diagrams, could place dot plots in Year 2 or connect with block diagrams when in Year 3)
Using aids and tools (technology) Y1-2	[Link with Computing] 'Unplugged' activities and use of digital technology Use 'unplugged' computing ideas to support their data-related work, e.g. students can stand in a line or circle and then rearrange themselves based on a given sorting criterion such as height or birth month. [Link with England's Computing PoS, KS1 subject content] Pupils should be taught to use technology purposefully to create, organise, store, manipulate and retrieve digital content Why? To provide statistically meaningful embodied activities.

At Key Stage 2, in addition to the enquiry cycle, the increasing mathematical sophistication/processes and availability of digital tools provide access to a growing set of statistical methods and representations.

During Years 3 and 4

England's NC statement	Year 3: Pupils should be taught to: • interpret and present data using bar charts, pictograms and tables • solve one-step and two-step questions [for example, 'How many more?' and 'How many fewer?'] using information presented in scaled bar charts and pictograms and tables. NSG: Pupils understand and use simple scales (for example, 2, 5, 10 units per cm) in pictograms and bar charts with increasing accuracy. They continue to interpret data presented in many contexts. Year 4: Pupils should be taught to: • interpret and present discrete and continuous data using appropriate graphical methods, including bar charts and time graphs. • solve comparison, sum and difference problems using information presented in bar charts, pictograms, tables and other graphs. NSG: Pupils understand and use a greater range of scales in their representations. Pupils begin to relate the graphical representation of data to recording change over time.
Statistical enquiry Y3-4 (and sub-strand on Posing questions)	[Link with Science] The Years 3-4 strand on 'Working scientifically' in England's Science NC includes 'asking relevant questions and using different types of scientific enquiries to answer them', 'setting up simple practical enquiries, comparative and fair tests', and 'gathering, recording, classifying and presenting data in a variety of ways to help in answering questions' 'gathering and recording data to help in answering questions.' Create and conduct investigations for different purposes (e.g. solving a problem, making a decision) using the statistical enquiry cycle to understand issues relevant to their immediate locality and context, using primary data they have collected. (JJ) Year 3: Investigating questions where the response is a count or a discrete measurement (i.e. a whole number), e.g. How many teeth have been lost by the students in our class? What are the shoe sizes in the class? (adapted from NZ) Year 4: Investigations that use numerical data which may need rounding to an appropriate unit or part of a unit, based on the context, e.g. How long does it take us to run 400 m? (adapted from NZ) Posing questions: Ask Questions That Data Can Answer

	Years 3-4: Students could start with asking 'who is best in our class at shooting/saving penalties?' and develop that to devise a question/trial that accounts for variability (potentially leading to emergent/informal ideas on samples and averages). Why? The statistics education literature (e.g. GAISE II) underscore the importance of situating statistical learning in meaningful statistical investigations.
Data Stewardship Y3-4	[Link with PSHE] When collecting and presenting data, include information on the purpose, origin and accuracy of data. Understand that data can show some things but not others. (DE4E) [Link with PSHE] Describe simple methods of processing data securely e.g. choosing and setting a secure password. (JJ) Understand that people have rights about how their data is used. (JJ) [Link to England's Computing PoS, KS2 subject content] Pupils should be taught to use search technologies effectively, appreciate how results are selected and ranked, and be discerning in evaluating digital content; use technology safely, respectfully and responsibly; recognise acceptable/unacceptable behaviour; identify a range of ways to report concerns about content and contact. [Link to England's PSHE KS2 PoS, Media literacy & digital resilience] Pupils learn: L12. how to assess the reliability of sources of information online; and how to make safe, reliable choices from search results Why? Building a progression for metadata that clarifies provenance of data and would support data stewardship (and later work on recognising and addressing bias).
Handling Data Y3-4	Collect and record numerical data that is relevant to their immediate locality and context, rounding with increasing accuracy, e.g. 2, 5 or 10 units per cm, or grouping appropriately, e.g. number of teeth lost, how many skips in 30 seconds. Recognise that a numerical variable can arise from measuring or counting, and a categorical variable can arise from classifying and/or sorting (adapted from NZ) Why? Present across national curricula, with NZ emphasising more formal language and England having statements for visual representations, e.g. Year 3: Statistics Notes and guidance (non-statutory) Students understand and use simple scales (for example, 2, 5, 10 units per cm) in pictograms and bar charts with increasing accuracy.

Statistical thinking and methods Y3-4	Collect and record multiple measures of the same numerical attribute using units based on standard units, e.g. use 5ml spoons to measure capacity, and describe the variability. [Link to measurement] Ask and answer questions about a numerical or categorical data set, using variable names and identifying the operations required for one-step and two-step comparison, sum and difference problems. Distinguish between when to use a particular value or the frequency for a given value (NZ). Why? Measuring as a source of numerical data, where methods are applied to real, meaningful contexts, appears in Y3 in NZ or Y4 in Singapore (P3 work in groups to measure). NB Singapore restrict to solving only one-step problems in P3/4 (which is our Y4/5). Link to Oracy – communicating the meaning of numbers both in the context and in the calculation (value, frequency, comparison).
Probabilistic Reasoning Y3-4	[Link with Oracy] Make predictions with evidence and caution, using language of uncertainty (likely, unlikely, certain, impossible, may be, because) and names of relevant variables. [Link with Citizenship] Use chance experiments for fair choices Why? All sources defer formal probability until KS2 at earliest and the underpinning work with fractions is still developing in years 3 & 4. Scotland, DE4E, GAISE11 emphasise early attention to expressing uncertainty when reasoning, and starting to associate likelihood with frequency (i.e. prediction with evidence). DE4E, Scotland introduce purposeful experiences with chance experiments.
Data representation Y3-4	Interpret and present data using bar charts, pictograms and tables Solve one-step and two-step questions [for example, 'How many more?' and 'How many fewer?'] using information presented in scaled bar charts and pictograms and tables. (England's NC) Recognise that a good data visualisation includes, where appropriate: - a title that gives the purpose of the visualisation - variable(s) (e.g. labelled on the axis) - the group the data is from - units for a numerical variable - values or categories - frequency, with the scale starting at 0. (NZ) Describe the data clearly by identifying any trends or patterns found, using descriptive language and terms such as "most," "least," "greater than," "less than," and "equal to." (DE4E) Why? Data representations are fairly consistent across various national Primary curricula, though with differences in year groups when particular representations are introduced. NZ does more and earlier, Singapore has 'with scales' in Primary Two'; bar graphs in Primary Three; tables in Primary Four.

	Though evidence for particular choices unclear. Given that students are already undertaking work on drawing and using charts and tables, it makes sense to include guidelines for clarity of presentations.
Using aids and tools (technology)	[Links with Computing and Science] Use of data-loggers Use simple paper-based and digital forms, and equipment such as scales, timers and data-loggers, to gather data for a **purpose relevant** to their immediate locality/ context. (adapted from JJ) Use calculators to perform calculations with small sets of data gathered for a **purpose relevant** to their immediate locality/ context. (JJ) [Link to England's Computing PoS, KS2 subject content] Pupils should be taught to select, use and combine a variety of software (including internet services) on a range of digital devices to design and create a range of programs, systems and content that accomplish given goals, including collecting, analysing, evaluating and presenting data and information Why? Technology is an essential underpinning for data science and needs to be integrated into students' experience of data education.

During Years 5 and 6

England's NC statement	Year 5: Pupils should be taught to: • solve comparison, sum and difference problems using information presented in a line graph • complete, read and interpret information in tables, including timetables. NSG: Pupils connect their work on coordinates and scales to their interpretation of time graphs. They begin to decide which representations of data are most appropriate and why. Year 6: Pupils should be taught to: • interpret and construct pie charts and line graphs and use these to solve problems • calculate and interpret the mean as an average. NSG: Pupils connect their work on angles, fractions and percentages to the interpretation of pie charts. Pupils both encounter and draw graphs relating two variables, arising from their own enquiry and in other subjects. They should connect conversion from kilometres to miles in measurement to its graphical representation. Pupils know when it is appropriate to find the mean of a data set.
Statistical enquiry Y5-6 (and sub-strand on Posing questions)	[Link with Science] The Years 5-6 strand on 'Working scientifically' in England's Science NC includes 'planning different types of scientific enquiries to answer questions, including recognising and controlling variables where necessary'. Create and conduct investigations for different purposes (e.g. solving a problem, making a decision, observing changes, better understanding of an issue/ risk) using the statistical enquiry cycle to understand issues **relevant** to their immediate locality and context, using primary data they have collected. (JJ) Contexts can include health, civics, digital interactions. Year 5: Investigate questions involving bivariate data with two categorical variables, e.g. what students in our class do at lunch time, and their gender. (adapted from NZ) Year 6: Investigate questions involving time-series data, e.g. how the mass of a kilogram of carrots varies over 5 days (or the height of a plant over time) (adapted from NZ) Posing questions: Ask Questions That Data Can Answer When students wonder about something (like "Why do some plants grow faster?"), help them turn it into a measurable question: "Do plants with more sunlight grow taller in two weeks than plants with less sunlight?" (DE4E) Why? The statistics education literature (e.g. GAISE II) underscores the importance of situating statistical learning in meaningful statistical investigations.

Data Stewardship Y5-6	[Links to England's PSHE PoS, Media literacy & digital resilience] Pupils learn … L12. how to assess the reliability of sources of information online; and how to make safe, reliable choices from search results L13. about some of the different ways information and data is shared and used online, including for commercial purposes L14. about how information on the internet is ranked, selected and targeted at specific individuals and groups; that connected devices can share information L15. recognise things appropriate to share and things that should not be shared on social media; rules surrounding distribution of images L16. about how text and images in the media and on social media can be manipulated or invented; strategies to evaluate the reliability of sources and identify misinformation. [For Link with Computing, refer to KS2 content in Y3-4 table] Consider the context, scope, and purpose of data in order to select variables of interest while recognizing those selections will have limitations. (adapted from DE4E) Recognize that some biases in data are neutral, while others can be harmful when making decisions and inferences, and some may not cause harm at all. e.g., neutral, preference of apples over oranges in a fruit study; harmful, surveying the coding club to generalize about all students. (DE4E) Identify how data collection can create risks (e.g., medical information, location, privacy, exclusion) for individuals or groups, and describe ways to protect personal information. (DE4E). Understand that apps on digital devices can leave data trails. (JJ) Why? Given potential statistical investigations in science or geography that relate to health, the environment, sustainability, there is the scope to include the above statements from DE4E, particularly as they seed or develop progressions for bias, risks, privacy, ethics and context. There is a progression in the statements above to do with developing complexity of investigations rather than evidence for year-by-year progression.
Handling Data Y5-6	Gather, clean, organise and store discrete and continuous primary data of relevance to local/ familiar issues and contexts using a variety of appropriate methods, e.g. simple paper and/or digital surveys/ forms; simple databases. (JJ) Describe the difference between primary and secondary data. (JJ) Collect and record bivariate (paired) and multivariate categorical data that is relevant to their immediate locality and context, e.g. height, birth season and favourite sport, or plant height and time. (NZ example: what students in our class do at lunch time, and their gender) Understand that measurements for continuous numerical data need to have the place value for rounding specified, e.g. to the nearest centimetre and

	create tables for continuous numerical data, using groupings, e.g. 0–0.99, 1–1.99, 2–2.99 (NZ) Additional at Year 6: Bivariate data includes time-series data with two numerical values, one time-based (NZ), e.g. height of plant over time. Explore how to record related or bivariate data to support visualisation and analysis e.g. plant growth data recorded by different students. Why? NZ covers this with bivariate data included at Year 5. England's NSG has students working with 'graphs relating two variables, arising from their own enquiry and in other subjects'. Existing work on data representation and statistical problem solving in England's curriculum could support the type of univariate and bivariate investigations suggested here. Note that time graphs represent observed data points rather than a functional relationship such as conversion from miles to km. Year-wise division are reflective of increasing complexity of measured data compared with counts.
Statistical thinking and methods Y5-6	Ask and answer simple questions identifying trends and comparing frequencies involving one or more variables. Use related data sets to describe and compare two groups using the language of shape (number of peaks, symmetry) most/least frequent categories, middle and spread. Use clustered bar charts to support analysis [Link to history , social geography] Year 6 Interpret and calculate the mean of a data set as an average or equal share. Why? Need to compare groups with data – gives meaning and purpose to summaries and shape. Singapore, NZ, GAISEII, DE4E all introduce between group comparisons early. JJ do have all three measures of central tendency here but neither NI, NZ or Singapore introduce mode or median in primary. NZ have both mean and range, but repeat in Year 7. Scotland delays mean until students can calculate and compare measures of central tendency approx Y 10-11.
Probabilistic Reasoning Y5-6	Make predictions about the outcome of a random selection in context using relevant data e.g. If a student in our class is selected at random, what type of music are they likely to prefer? Distinguish between outcomes that are equally likely and those that are more/less likely. [Link with number - follows from correspondence problems, currently Y3 & 4] Identify all the outcomes of a future event, for example tossing two coins, netball match.

	Why? JJ, DE4E and GAISEII emphasise using data for the purpose of prediction at this age. While NZ and Wales introduce the 0-1 probability scale and equally likely outcomes in KS2 , they revisit in years 7 and 8, suggesting repetition.
Data representation Y5-6	Year 5: Continuous numerical data can be organised in a table by grouping data into specific ranges of values. Paired categorical data can be visualised with a clustered bar graph (NZ) Year 6: Interpret and construct line graphs and use these to solve problems (England's NC with NSG: Pupils both encounter and draw graphs relating two variables, arising from their own enquiry and in other subjects.) In addition to the Years 3-4 attributes for good visualisations, have consistent width, spacing and use of colour for clustered bar graphs, with a key for the colours, and mention the source of the data. Choose and create appropriate data visualisations (using appropriate technology), and supporting examples, that reveal the narrative the data conveys. (adapted from DE4E) Why? The representations for paired categorical data at Year 5 (NZ) support meaningful investigations and sit between the bar charts done in Year 4 and the two-variable line graphs in Year 6. Pie charts are in current England's NC but can be removed: drawing them without technology should be left (if at all) for Year 7 or later when proportional reasoning is more secure. While pie charts may provide a context for undertaking proportional reasoning calculations, (a) the angular/circular representation differs from bar graph or scalar representations; (b) if only 'simple' values are being used (to give manageable angles) then the related artificiality should be clarified to students.
Using aids and tools (technology) Y5-6	[Links with Computing and Science] Use of unplugged activities, data-loggers, computers and other digital devices, software, apps, digital environments, and AI and Machine Learning related tools and approaches. Use calculators or computers to perform calculations with sets of data gathered for a **purpose relevant** to their immediate locality/ context. Use simple paper-based and digital forms, and equipment such as scales, timers and data-loggers, to gather data for a **purpose** relevant to their immediate locality/ context. (adapted from JJ) Use computers to generate data representations and visualisations. Use unplugged activities such as ProDaBi's Nutrition 'Data cards' to introduce principles of Machine Learning (ML) to answer questions such as 'how can we determine whether a snack is healthy or not based on its attributes?' The activity introduces data as a model of the world, data processing, exploration

	of data, and use of data-driven ML (via decision trees) for making predictions and decisions. See Appendix 2. Use no-code digital environments such as *Somekone*, a social media simulator and explainable AI tool that is designed to help learners understand some of the fundamental processes behind social media platforms, including data collection, engagement, profiling, recommendations, and filter bubbles. https://www.raspberrypi.org/blog/how-ai-shapes-your-feed-an-explainable-social-media-simulator-for-the-classroom/ [Link to England's Computing PoS, KS2 subject content] Pupils should be taught to select, use and combine a variety of software (including internet services) on a range of digital devices to design and create a range of programs, systems and content that accomplish given goals, including collecting, analysing, evaluating and presenting data and information Why? Technology is an essential underpinning for data science and needs to be integrated into students' experience of data education. Also, certain data (given its size, format, complexity or dynamic nature) can only be accessed using technology.

Secondary

In addition to the continuing and overarching presence of the statistical enquiry cycle, the growing and powerful mathematical toolkit enables an increase in (and a formalisation of) the statistical methods and representations available to students. Probability, the types of variables, measures of central tendency and spread, and notions of populations, samples and distributions are formalised. These, and a wider range of representations, enable attending to variation, relationships and increasing complexity with greater precision. Students are able to strengthen their problem-solving, analysis, inferences, predictions and communication, supporting their statistical sense-making when there is too much or too little data. Where/when available, technology enables working with large, varied, complex and dynamic data sets.

During Years 7, 8 and 9

Note: There is no mention of statistics or data in the introductory 'Purpose of study' and Aims in England's KS3 Mathematics PoS. 'Working mathematically' touches on 'move freely between … representations' within 'Develop fluency' but examples are not statistical. 'Use language and properties precisely to analyse …' does include probability and statistics.

"Reason mathematically" does finally have a statement on statistical exploration, inference and argumentation.

Also, the Statistics content is largely in terms of representation and interpretation, and mainly lies in the Data representation section.

England's NC statement	KS3: Pupils should be taught to: • describe, interpret and compare observed distributions of a single variable through: appropriate graphical representation involving discrete, continuous and grouped data; and appropriate measures of central tendency (mean, mode, median) and spread (range, consideration of outliers) • construct and interpret appropriate tables, charts, and diagrams, including frequency tables, bar charts, pie charts, and pictograms for categorical data, and vertical line (or bar) charts for ungrouped and grouped numerical data • describe simple mathematical relationships between two variables (bivariate data) in observational and experimental contexts and illustrate using scatter graphs. **Probability** Pupils should be taught to: • record, describe and analyse the frequency of outcomes of simple probability experiments involving randomness, fairness, equally and

	unequally likely outcomes, using appropriate language and the 0-1 probability scale • understand that the probabilities of all possible outcomes sum to 1 • enumerate sets and unions/intersections of sets systematically, using tables, grids and Venn diagrams • generate theoretical sample spaces for single and combined events with equally likely, mutually exclusive outcomes and use these to calculate theoretical probabilities.
Statistical enquiry Y7-9 (and sub-strand on Posing questions)	[Link with Science] The KS3 strand on 'Working scientifically' in England's Science NC has guidance on conducting scientific enquiry and use of data. [Link with Geography] The Geography NC KS3 'Geographical skills and fieldwork' strand includes 'use fieldwork in contrasting locations to collect, analyse and draw conclusions from geographical data, using multiple sources of increasingly complex information.' Create and conduct observational and experimental investigations for different purposes (e.g. solving a problem, making a decision, observing changes, better understanding of an issue/ risk) using the statistical enquiry cycle to understand and propose solutions to issues **relevant** to their immediate and wider locality and context, combining primary and secondary data. (JJ) Year 7,8: Investigate statistical questions that clearly identify the variable, group of interest, and the intent of an investigation such as: are our feet the same length? (adapted from NZ) Year 9: Plan and collect multivariate data to respond to a statistical question where at least one variable is categorical and at least one is numerical (adapted from NZ) Posing questions: Ask Questions That Data Can Answer Years 7-9: Give students a broad topic they care about (video games, sports, music) and challenge them to write three different questions: one they could answer by counting, one by comparing groups, and one that would require collecting information over time. (DE4E) Why? The statistics education literature (e.g. GAISE II) underscore the importance of situating statistical learning in meaningful statistical investigations.
Data Stewardship Y7-9	[Links to England's PSHE PoS, Media literacy & digital resilience] Pupils learn … L20. that features of the internet can amplify risks and opportunities, e.g. speed and scale of information sharing, blurred public and private boundaries and a perception of anonymity L27. to respond appropriately when things go wrong online, including confidently accessing support, reporting to authorities and platforms]

	Acknowledge that options and choices are available for data collected about individuals, and recognize that what is gathered or excluded can have consequences. (DE4E) Recognise the purposes for which data are collected and processed, and attend to privacy, security, potential for bias and unfair representation in a responsible manner. (OU, building on GAISE II) Recognise that some data might be sensitive and may need to be processed differently from other data. Understand different ways to protect data including personal data on digital devices. Know that ethical data collection involves seeking consent, transparency and limiting harm to those from whom data is collected. (JJ) Critically consider how secondary data has been gathered, including potential biases. (JJ) Ask questions regarding the origins of data and any specified or automated measures (adapted from DE4E) Specify ways that data provide incomplete information relative to the object being studied. (DE4E) Identify how biases in data affect inferences and questions. (DE4E) [Link with history] Examine historical examples of harmful data practices to inform ethical data use. (DE4E) [Link to England's Computing PoS, KS3 subject content] Pupils should be taught to create, re-use, revise and re-purpose digital artefacts for a given audience, with attention to trustworthiness, design and usability; understand a range of ways to use technology safely, respectfully, responsibly and securely, including protecting their online identity and privacy; recognise inappropriate content, contact and conduct and know how to report concerns. Why? Statements from JJ and from DE4E develop data protection and bias progressions respectively (from KS2). DE4E goes further but that may be somewhat advanced given where we are (e.g. Design data collection methods that address privacy, consent, and fair representation of different groups.)
Handling Data Y7-9	Gather, organise, clean, and store (primary and secondary) data to understand issues relevant to their immediate and wider locality and context, using appropriate methods including digital technologies. (JJ) Distinguish between types of variables (e.g., categorical or quantitative); engage with data arising from familiar sources (e.g. mobile phones) and from multiple sources. (OU, building on GAISE II) Use appropriate data handling methods depending on the type of data or setting (e.g., a survey or experiment); devise appropriate data collection plans (OU, building on GAISE II)

	Connect related datasets; structure data so that it can be efficiently searched, manipulated and analysed (using technology) Why? Progression based on increasing complexity (could build up to it by spreading over year groups)To make explicit handling data competencies not delineated in England's NC.
Statistical thinking and methods Y7-9	Ask and answer statistical questions that clearly identify the variable, group of interest, and the intent of an investigation. Describe simple mathematical relationships between two variables (bivariate data) in observational and experimental contexts and illustrate using scatter graphs. Identify outliers as data points that are much bigger or smaller than most data points and consider reasons for removing or keeping these. Identify and use the median as the middle value of sorted numerical data e.g. in a dot plot or stem and leaf diagram, and the range as a measure of spread. Interpret and use measures of central tendency (median, mode, mean) and spread (minimum, maximum, range, consideration of outliers) to make sense of ungrouped data relevant to their immediate and wider locality and context based on the purpose for which the data were collected. Identify when a data visualisation cannot be interpreted accurately due to missing information such as scale or variable name. Why? Two strands of analysis separate – those used for categorical data and those with bivariate numerical data. GAISEII emphasizes familiarity with median before mean to represent central tendency as it follows from comparison problems and from 'Distinguish between when to use a particular value or the frequency for a given value' in Y5/6. Scotland delays using all three measures until Y10,11 when their use can be compared. NZ, GAISEII recommend an ongoing emphasis on comparing two groups in context, now using a wider range of summary statistics.
Probabilistic Reasoning Y7-9	[Link to proportional reasoning means that, in England's curriculum, the following could all be year 9]. Use probabilities to determine how many times I expect an outcome or event to occur. Communicate the probabilities of events in context (e.g. health outcomes) using natural frequencies and probabilities. Generate theoretical sample spaces for single experiments with equally likely, mutually exclusive outcomes and use these to calculate theoretical probabilities for outcomes and events. Understand that the probabilities of all possible outcomes sum to 1.

	Record, describe and analyse the frequency of outcomes of simple chance experiments involving equally likely and unequally likely outcomes e.g. spinner, selecting a person from a panel. Use results to calculate estimated probabilities of future outcomes and events. Understand that sets of repeated trials from the same experiment may vary. Extend chance experiments to at least 100 trials (aggregating results or using a simulation) and compare the experimental probability of an outcome to its theoretical probability. Why? The 0-1 probability scale depends on proportional reasoning which is the focus of Yrs 7-9. NZ and Scotland introduce theoretical probability in year 7, Singapore in Y9 and for single events only. RSS and JJ recommend limiting use of dice, cards and coins in favour of random selections from data that have meaning and purpose (i.e. using natural frequencies rather than abstract 'fairness' as the basis for theoretical probabilities). DE4E, NZ emphasise sampling as a means of solving prediction problems. RSS, DE4E, NZ, Scotland all emphasise that lower secondary students should understand that repeated sampling reduces the effects of random chance and the number of trials experienced should be sufficient to demonstrate this. Singapore restricts sample spaces to single events. NB wording of experiment and event is clarified (sample space is not of an event).
Data representation Y7-9	Construct and interpret appropriate tables, charts, and diagrams, including frequency tables, bar charts, pie charts, and pictograms for categorical data, and vertical line (or bar) charts for ungrouped and grouped numerical data using technology where possible. Years 7 and 8 (NZ): Choosing and constructing an appropriate data visualisation according to the data type. Categorical data can be visualised through dot plots and bar graphs. Paired categorical variables can be visualised through a stacked bar graph or a clustered bar graph. Bivariate time-series data can be visualised through a time-series graph. A good data visualisation should allow viewers to discern the variable or variables and who the data was collected from, and then, depending on the type of visualisation, additional information such as units for numerical variables, frequency, proportions, patterns, and trends. (NZ, for Years 9 and 10) Recognize that data visualizations need explanations to tell their story. (DE4E) Interpret visualised data by identifying stories in the data and critically compare and evaluate the different ways visualisations communicate data to understand issues relevant to their immediate and wider locality and contexts and according to the **purpose** for which the data were collected. (JJ)

	Critically consider how secondary data has been represented, including looking for the possibility of miscommunicating or misrepresenting data. (JJ) Choose and generate visualisations of data gathered to understand issues **relevant** to their immediate and wider locality and context using appropriate technology according to the **purpose** for which the data were collected. (JJ) Present data in a way that is accessible and engaging, while considering the specific needs, interests, and knowledge level of the audience. (DE4E) Why? To elaborate on England's NC, fill in progressions arising in KS2 and bridge to KS4, and connect with Statistical Enquiry.
Using aids and tools (technology) Y7-9	[Links with Computing, Geography and Science] Use of data loggers, computers, GIS systems, software including spreadsheets, apps, other digital devices and environments, and AI and Machine Learning related tools and approaches. Use digital tools (e.g. online forms, data loggers) to gather data for a **purpose relevant** to their immediate and wider locality and context. Use spreadsheets and databases to store, organise, analyse and present data, and to generate visualisations of data. Write simple formulae in spreadsheets to perform calculations involving data gathered for a **purpose relevant** to their immediate or wider locality/ context, involving more than one cell. (JJ) Use computers to generate data representations and visualisations. Use digital environments such as *Somekone*, a social media simulator and explainable AI tool that is designed to help learners understand some of the fundamental processes behind social media platforms, including data collection, engagement, profiling, recommendations, and filter bubbles. https://www.raspberrypi.org/blog/how-ai-shapes-your-feed-an-explainable-social-media-simulator-for-the-classroom/ [Link with England's Computing PoS, KS3 subject content] Pupils should be taught to understand how data of various types (including text, sounds and pictures) can be represented and manipulated digitally, in the form of binary digits Why? Technology is an essential underpinning for data science and needs to be integrated into students' experience of data education. Also, certain data (given its size, format, complexity or dynamic nature) can only be accessed using technology.

During Years 10 and 11

As in the earlier instances, there is no mention of any statistics in the introductory information in the England KS4 mathematics PoS.

The aim of these years is for students to make reasoned claims about data, choosing amongst basic statistical representations and measures as evidence. They argue more precisely about likelihood, using probabilities, and start to make decisions that involve both the probability and the consequence of events. They can use these skills to critique claims in the media. They gain experience with comparing and generating samples that supports them with informal inference (i.e. making claims about unknown populations based on observed samples) and will underpin formal methods in any later study.

Despite mentioning of issues around samples and populations, there is hardly any mention of actual handling of data in England's KS3 and KS4 PoS. Handling data merits explicit inclusion to avoid presenting a distorted image that statistics is simply the mathematical analysis of already represented data.

England's NC statement	**Probability** In addition to consolidating subject content from key stage 3, pupils should be taught to: • apply the property that the probabilities of an exhaustive set of mutually exclusive events sum to one • use a probability model to predict the outcomes of future experiments; understand that empirical unbiased samples tend towards theoretical probability distributions, with increasing sample size • calculate the probability of independent and dependent combined events, including using tree diagrams and other representations, and know the underlying assumptions • **{calculate and interpret conditional probabilities through representation using expected frequencies with two-way tables, tree diagrams and Venn diagrams}.** **Statistics** In addition to consolidating subject content from key stage 3, pupils should be taught to: • infer properties of populations or distributions from a sample, whilst knowing the limitations of sampling • interpret and construct tables and line graphs for time series data • **{construct and interpret diagrams for grouped discrete data and continuous data, i.e. histograms with equal and unequal class**

	intervals and cumulative frequency graphs, and know their appropriate use} • interpret, analyse and compare the distributions of data sets from univariate empirical distributions through: • appropriate graphical representation involving discrete, continuous and grouped data, {**including box plots**} • appropriate measures of central tendency (including modal class) and spread {**including quartiles and inter-quartile range**} • apply statistics to describe a population • use and interpret scatter graphs of bivariate data; recognise correlation and know that it does not indicate causation; draw estimated lines of best fit; make predictions; interpolate and extrapolate apparent trends whilst knowing the dangers of so doing.
Statistical enquiry Y10-11 (and sub-strand on Posing questions)	Create and conduct for different purposes using the full mathematical/ statistical/computational enquiry cycle in domains related to health, personal finances, civics, including using data that has been gathered from familiar digital interactions (e.g. YouTube views), attending to the context and variability in the data. (JJ) Design and conduct a statistical investigation requiring multivariate data, a sample or census, and use of technology (adapted from NZ) Posing questions: Ask Questions That Data Can Answer - Present students with a local issue (school lunch satisfaction, parking problems, etc.) and have them identify what data would be needed to understand different perspectives on the issue. - Have students find a news article making a claim and identify what additional data would be needed to verify, challenge, or strengthen that claim. (DE4E) [Could also use items from BBC's More or Less radio programme] Why? The statistics education literature (e.g. GAISE II) underscore the importance of situating statistical learning in meaningful statistical investigations. [Link with Science] The England's Science GCSE curriculum includes 'develop understanding … through different types of scientific enquiry that help them to answer scientific questions about the world around them', 'develop their ability to evaluate claims based on science through critical analysis of the methodology, evidence and conclusions, both qualitatively and quantitatively'. Working scientifically includes the scientific method and 'evaluating risks both in practical science and the wider societal context, including perception of risk', 'recognising when to apply a knowledge of sampling techniques to ensure any samples collected are representative', 'evaluating methods and suggesting possible improvements and further investigations.', and a detailed section on 'Analysis and evaluation' which covers 'applying the cycle of collecting, presenting and analysing data'. [Link with Geography] 'The Subject aims and learning outcomes' include 'using maps and Geographical Information Systems (GIS) and in researching

	secondary evidence, including digital sources; and develop their competence in applying sound enquiry and investigative approaches to questions and hypotheses (study like a geographer)’. The ‘Progression statement’ includes ‘a greater stress on the multivariate nature of 'human-physical' relationships and interactions’. The ‘Scope of study’ has two key areas that link to data education. Use of data: ‘Data’ should include both qualitative and quantitative data and data from both primary and secondary sources: fieldwork data; GIS material; written and digital sources; visual and graphical sources; and numerical and statistical information. Using data should include its collection, interpretation and analysis, including the application of appropriate quantitative and statistical techniques (a list of required skills and techniques is given in the Appendix); it also includes the effective presentation, communication and evaluation of material. Formulating enquiry and argument: ‘The ability to identify questions and sequences of enquiry to write descriptively, analytically and critically, to communicate their ideas effectively, to develop an extended written argument, and to draw well-evidenced and informed conclusions’ Not done for this Rapid Review but within each of the strands, one could link to specific data skills for geography listed in the Appendix: Use of mathematics and statistics in geography (includes graphical, numerical and statistical skills).
Data Stewardship Y10-11	Recognise the **purposes** for which data are collected and processed, and attend to privacy, security, potential for bias and unfair representation in a responsible manner. (JJ) Ensure data that may contain sensitive information are handled appropriately; appreciate how to manage data so that it is accessible, usable, re-usable and well-documented (OU, building on GAISE II) Explain how data-based decisions are revisited as new evidence or societal needs emerge, e.g., blood pressure cut-off numbers, dietary guidance, medical benchmarks (DE4E) Evaluate claims derived from data by questioning how phenomena are measured, categorized, or represented. (DE4E) Recognise that decisions may not just be based on data and data analyses, e.g. role of economic, social and cultural factors. [Link to England's Computing PoS, KS3 subject content] understand how changes in technology affect safety, including new ways to protect their online privacy and identity, and how to identify and report a range of concerns. Why? Progressions for the purposes, origins and safeguards for data collection and use.

Handling Data Y10-11	Gather, organise, clean, store, display and interrogate data **purpose**fully (primary and secondary data); Distinguish between types of variables (e.g., categorical or quantitative) and understand what characteristic they represent; engage with data arising from familiar sources (e.g. mobile phones) and from multiple sources. (JJ) Appreciate the need for samples 'to make inferences about a population without a census'. (NZ) Appreciate the value of randomisation and sample size for better inferences. (adapted from NZ, which provides parameters for certain situations) Plan and collect multivariate data to respond to a statistical question using a sample or census (NZ) Have some experience of working with large data sets and know how to address missing data. (OU, building on GAISE II) Select data analysis techniques appropriate for the type of data being processed; connect data from different sources, including for large data sets. (OU, building on GAISE II) Why? Clarifying the handling of data left implicit or absent in England's NC.
Statistical thinking and methods Y10-11	Engage in exploratory data analysis: find, describe and analyse patterns, relationships and trends Describe covariation between two variables (JJ): use and interpret scatter graphs of bivariate data; recognise correlation and know that it does not indicate causation; draw estimated lines of best fit; make predictions; interpolate and extrapolate apparent trends whilst knowing the dangers of so doing. (England's NC) Compare conditional frequencies across categorical variables (linked to proportional reasoning) (JJ) Anticipate, recognise and account for variability in data, including via error or chance, [and in sampling] and appreciate how it shapes analyses and predictions. Explain how removing or including datapoints (e.g. outliners) affects the range, median and mean. {**Estimate medians and means for grouped data sets**}. Summarise and compare the distributions of univariate, empirical data sets through appropriate features and measures of distributions such as gaps, symmetry, mean, median, range, five-number summaries. Make informal inferences e.g. by using a 75% to 50% comparison rule. Why? All sources agree by year 11, students should make choices and show precision in exploratory data analysis by using range of measures of central tendency and spread as a means of summarising and comparing groups with data. This includes comparing and understanding the effect in context of

	different measures. Understanding of variability extends to understanding the effect on measures when data is added or removed from the set. Five number summaries organise measures already learnt and connect with boxplots to support comparisons between groups. NZ have added detail by specifying and naming a '75% to 50% comparison rule' for informal inference (if more than 50% of group *B*'s data (i.e. the median) is larger than 75% of group *A*'s data, then we can claim that B tends to take larger values than A). Trends in bivariate data can be connected to functional relations in algebra although there should be a continued focus on describing the variability of statistical data (i.e. not reducing the data to the model).
Probabilistic Reasoning Y10-11	[Link with financial literacy] Communicate the risk of events in context: make informed choices based on probability and consequences e.g. decide between flood defence systems based on cost and probability. Interpret contextual statements and diagrams to construct sample spaces for combined/joint experiments, recognising when the sample space has equally likely outcomes (e.g. rolling two fair dice) and when outcome probabilities differ (e.g. true/false results on two medical tests), and use these representations to calculate probabilities. Decide whether simple combined/joint events are independent or dependent. Calculate their probabilities including using tree diagrams and other representations. {**Calculate and interpret conditional probabilities using expected frequencies with two-way tables, tree diagrams and Venn diagrams.**} Identify situations when distinguishing from random chance is important e.g. medical trials. Use probabilistic reasoning to make sense of outliers or other unusual/ unexpected features of data. Estimate the probability of an event by carrying out chance experiments repeatedly using a simulation, and decide a criterion for stopping the trials and communicate the probability {**as an interval**} e.g. the estimated probability of two wins, one loss is 0.4 **{between 0.36 and 0.41}.** Why? RSS recommends limiting use of dice, cards and coins in favour of predictions from natural frequencies in real world contextual data. Students should not only calculate risk but know ways of communicating it. DE4E, Scotland address critical data thinking (and financial literacy) by considering consequences as well as probabilities, i.e. starting to assess the risk of a limited range of outcomes. Systematic consideration of sample spaces allows consideration of the assumptions underlying equally likely and unequally likely outcomes.

	DE4E, NZ, GAISEII emphasise random sampling and simulation as processes to solve prediction problems and to understand the effects of random variability and expect increasing awareness of the effect of decisions taken. NB curriculum uses "combined" events and probabilities at GCSE but "joint" events at A level and beyond.
Data representation Y10-11	Interpret and construct tables and line graphs for time series data **{Construct and interpret diagrams for grouped discrete data and continuous data, i.e. histograms with equal and unequal class intervals and cumulative frequency graphs, and know their appropriate use}** [Use] appropriate graphical representation involving discrete, continuous and grouped data, {including box plots} (Above competencies are from England's NC. Box plots are in bold (not required for all) in England's NC but we suggest including them for all – see 'Why?' below) Use and interpret scatter graphs of bivariate data Draw estimated lines of best fit (England's NC) Choose and, importantly, compare, **connect** and combine different data representations, generated with appropriate technology; use appropriate representations, (e.g. proportions for categorical data, two way tables, tree diagrams and confusion matrices); represent the variability in data using appropriate visualisations. (JJ) A good data visualisation should allow viewers to discern the variable(s) and who the data was collected from, and then, depending on the type of visualisation, additional information such as frequency, proportions, patterns or trends, and units for numerical variables. Select appropriate scales for data (NZ) Use data to explain trends and predict future outcomes based on those trends. (DE4E) Identify an audience of interest, and tailor data stories to that audience (DE4E) Why? To elaborate on England's NC, fill in progressions arising in KS2 and bridge to KS4, and connect with Statistical Enquiry. Regarding box plots, they are useful for comparing two groups and schools working with Realistic Mathematics Education (RME, see www.rme.org.uk) have found that they are accessible for students in Year 9 and above. Some students can confuse the length of each of the four sections of a box plot with the related frequency, `but it can be clarified that each section constitutes a quarter of the data set, which helps connect the box plot with the shape of the underlying distribution. JJ also provide a reference for the use of box plots at KS3 (Saldanha & Hatfield, 2021).

Using aids and tools (technology) Y10-11	[Links with Computing, Geography and Science] Use of data loggers, computers, GIS systems, software including spreadsheets, apps, other digital devices and environments, and AI and Machine Learning related tools and approaches. Use a range of **relevant** computational tools, including calculators, spreadsheets, mathematical and statistical software, data analysis packages, visualisation tools, and machine learning algorithms. (JJ) Use computers to process, connect, and analyse data and generate representations and visualisations. Select a no-code, low-code or high-code digital tool that is suited for the intended task. (DE4E) Use coding-based activities such as ProDaBi's digital environment (includes CODAP and use of Python) which allows students to explore social media platforms' use of online gaming behaviours to decide which advertisement is shown to a player, to introduce principles of Machine Learning (ML) and computer-based implementations of data-driven ML and decision trees. See Appendix 2. [Link to England's Computing PoS, KS4 subject content] Pupils should be taught to develop and apply their analytic, problem-solving, design, and computational thinking skills Why? Technology is an essential underpinning for data science and needs to be integrated into students' experience of data education. Also, certain data (given its size, format, complexity or dynamic nature) can only be accessed using technology.

5. Implications for the current revision of the National Curriculum in England

The Curriculum and Assessment Review (DfE, 2025a) has recommended an evolution of the existing curriculum in England. This is unlikely to meet the vision of The Royal Society's Mathematical Futures report, but it acknowledges the strengths of much of our mathematics teaching and the demands on teachers of rapid change. There are specific areas in which the government's response to the review (DfE, 2025b) envisages making relevant changes. While not named as such, data education is central to the expectation that "The refreshed curriculum will teach pupils AI literacy, how to use data to complete tasks and solve problems, as well as learning about bias in technology" (DfE 2025b, p24). The government also resolves to tackle the challenge of equipping young people for a changing world by emphasising "the value of secure knowledge, the process of questioning and critical enquiry and weighing up evidence across information and sources" (ibid, p23). Again, this vision is central to data education and cannot be achieved without enabling students to reason mathematically and statistically about issues that are relevant to them.

Coherence

One task given to the curriculum drafters is to create a connected online version of the curriculum which visually represents the links within and between subject areas. In the tables above, we have included the main links between mathematics and other subjects that are present in our source curriculum documents. Notably this includes science, citizenship (for which there will be a new statutory requirement to teach in KS 1 and 2), computing (in which "pupils will know from a young age how computers can be trained using data" (ibid, p29), and oracy (for which there will be a cross-curricular framework).

These links may be used to distribute the data education competencies across subject areas and they may equally be used to structure a holistic enquiry that attends to all elements of data education relevant to that age range. This coherence is of particular relevance in primary years, where one teacher is responsible for many subjects. In secondary, the links may serve as a resource to plan schemes of work that are coherent across subjects and importantly include the relevant mathematical and statistical foundations. Pittard's (2018) analysis of data science skills within the existing national curriculum warned that, while opportunities to use data were implied in the non-statutory guidance of both mathematics and science, these connections are not made explicit and may depend on the practices of individual teachers and schools. This review is an opportunity to provide such coherence for all students.

One of the overarching principles is that students need to engage in the full investigative cycle, including posing and refining questions in context, leading to obtaining, evaluating, refining and communicating solutions, and involving iteration as necessary. Projects that span mathematics and science, mathematics and geography, or mathematics and computing could provide opportunities to develop such skills but there is no support in England's secondary school system for such collaborations to be embedded at scale, and their availability to all learners in years 10 and 11 would depend on subject choice. It seems important then that experiences that represent the full statistical enquiry cycle are embedded in mathematics, including some considerations of data stewardship.

Specificity

Curriculum drafters have to produce a PoS that is manageable in terms of teaching time. The DfE has highlighted specificity as a particular concern in creating a curriculum that works for all students: "A lack of specificity can hinder effective sequencing and lead to variation in what is taught, which – in turn – hampers good transitions between key stages. It can also lead to overloading of the curriculum as teachers try to cover everything in equal depth" (DfE, 2025b, p13). They emphasise that "**we will identify where subject-specific disciplinary skills are not clearly described or not emphasised adequately, and ensure these are properly reflected in those subjects' programmes of study**"(p23, bold in original).

Our synthesis of the source documents thus supports a move to greater specificity. We argue that this greater specificity is necessary to cover the breadth of data education and because the current curriculum content in statistics does not make explicit the most important progressions in statistical and probabilistic reasoning. As indicated above, the current statements typically specify key actions (describe, interpret and compare) on types of data and data representations. They do not clarify progression in: how data should be handled, relevant contexts, the kind of purposeful questions - involving description, prediction, variability and inference - that data allows you to answer, the levels of uncertainty in those answers and the statistical thinking that manages this uncertainty. Although many schools will understand the background of these compact and limited statements, they do not emphasise adequately the range of data skills or an appreciation of statistical thinking.

To give one example, in year 6, currency conversion graphs are offered as an example of statistical line graphs, whereas they offer no variability – their conversion rate is not random or observed but fixed. Instead, they are an example of functional covariation: the variables change according to a fixed and predictable relationship. Thus, a statistical representation, that should be used to find trends amongst variable data, is being put at the service of mathematics, with its statistical purposes omitted or left unclear to teachers and students.

Specifically, we recommend:

- Expand the explicit statistical content within mathematics to include statistical enquiry and the creation, collection and handling of data. Statistics is more than mathematical operations on numbers, data or representations.
- Include a statement about the nature of statistics in the Aims and Purposes section of the Mathematics POS, such as: "Statistics focuses on data and uncertainty. It develops students' understanding of how to collect, organise, and interpret data in context, and how statistical thinking supports informed decision making" (taken from the NZ curriculum)
- Ensure that all examples and guidance in the statistics section address statistical thinking and cannot be taught or assessed purely as an exercise in mathematics.

- Specify that students from Year 5 onwards should compare groups using related data sets. This gives a purpose to describing the shape of distributions and, later, their summary measures and also provides experiences of variability that underpin inference. Consider including an explicit rule-of-thumb for informal inference at GCSE level, for example see pages 42-43 above for the 75% to 50% rule. Consider also teaching the interpretation of boxplots for all Year 10-11 students as these are routinely used for group comparisons.
- Specify that students of all ages demonstrate probabilistic reasoning by making predictions *from evidence* and *with caution*. There is currently no explicit probability in primary mathematics, but students will be using concepts of likelihood in a range of contexts and should be encouraged to connect these with frequency. In later years, probabilistic reasoning contributes to media literacy by giving students tools to consider natural frequencies and to identify whether 'dramatic headline' events are common or rare. GCSE students should be asked to assess risk by making decisions that consider both probability and impact.

Demand on students and teachers

Explicit mention of authentic and meaningful statistical practice may increase word count of the mathematics PoS but, as desired by the government's principle of greater specificity, it need not imply a measurable increase in curricular time if the students are already undertaking statistical investigations. The expanded presentation of data and statistical elements serves more to raise awareness of holistic issues critical for data and statistical education and to support the sequencing and progression of learning.

Sequencing and progression

The government's response to the curriculum review (DfE, 2025b) has identified a need for greater attention to sequencing in the curriculum. This is particularly important for KS3 transitions where better sequencing is recommended for "help[ing] teachers to plan lessons so that they are not repetitive, disengaging or overly focussed on what they will be assessed on at key stage 4" (ibid, p13). An additional gain of introducing greater specificity in the curriculum is to demonstrate how statistical competencies, and even the overarching activity of statistical enquiry, should develop every year. Students' increased knowledge and fluency with mathematical and statistical tools, and with contexts and questions in which data is useful, allow a deeper understanding of how statistics approaches and models questions of variability and uncertainty.

It is notable in several source curricula (NZ, Singapore) that statistical measures such as mean that are taught in the final years of primary reappear, not necessarily with greater depth, in the lower secondary curriculum. While there is no research evidence about the ideal order of teaching, it seems plausible that some of these concepts are being taught before the relevant mathematical foundations are secure.

Proportional reasoning is important for data education. For data analysis, the magnitude and meaning of numerical information are usually understood through relative frequencies: the question 'How many?' is less interesting than 'How many out of how many?' or 'How many for each?' Most sources recommend that initial measures of probability are grounded in observed frequencies from chance experiments, again requiring a proportion.

As examples of helpful sequencing, we recommend that:

- Mean is taught when students have a secure foundation in equal sharing.
- Probability measure on the 0-1 scale is taught when students have a secure foundation in proportional reasoning.
- In KS2 and 3, students are encouraged to use multiple methods for describing the shape, central tendency and spread of a data set. For example, identifying a minimum and maximum should be as acceptable as calculating a range.
- Pie charts are removed from year 6 (currently included because they offer an opportunity to practice finding fractions and percentages of 360), with interpretation of pie charts introduced only when proportional reasoning and percentages are more secure. Note there is longstanding advice from data visualisation that pie charts are poor data representations (e.g. Cleveland and McGill ref 1984).
- Consider removing the construction of bar graphs, histograms and (if retained) pie charts by hand, in favour of using technology.

Digital methods

The government's response (DfE, 2025b) states that "in some subjects, digital methods now influence the content and how it is taught. We will work with experts to assess the validity of digital practice in these subjects, the evidence of whether this can be done robustly and whether it merits inclusion in the new curriculum. Where it does, **we will include a requirement for the relevant digital content in those subjects' programmes of study and we will ensure that it**

aligns with the computing curriculum, to reduce the risk of duplication. (p24, bold in original)

Given our sense that digital technologies are integral to DE, this offers an opportunity to ensure better integration of digital technologies in Mathematical Data Education (MDE).

Implications in the longer term

As touched on earlier in this section, the current revision of the National Curriculum in England should be a step towards the JMC/RS vision for MDE in England. Though the limited and rapid revision means that much will still need to be done. There is a nod to students learning about data-driven computing and AI. Even if DE (or MDE) is successful in obtaining a curricular 'requirement' for digital content, many of the barriers to integrating digital technologies in mathematics and data education, outlined in Crisan et al. (2023), will require substantial work stretching into the longer term.

There is still little evidence and no clear conclusion on how best DE sits in the curriculum, whether stand alone, within an existing subject or cross-curricular. If DE is to sit within mathematics in the classroom, then mathematics teaching will have to do justice to the overarching principles set out in Section 3. This raises a host of issues relating to implementation. Data Education is variably implemented in the nations of the UK. While Scotland has a strong DE curriculum and teaching resources, all UK nations face challenges relating to DE teaching capacity, teacher professional development, integration of technology, the enacted curriculum, and related research.

References

ACME PEY Expert Panel. (2025). *Data Education – Progression in the primary years.* Available from RS ACME (The Royal Society Advisory Committee on Mathematics Education). https://royalsociety.org/news-resources/projects/mathematics-education/royal-society-acme/

Bargagliotti, A., Franklin C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). 'Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report II'. American Statistical Association and National Council of Teachers of Mathematics. https://hdsr.mitpress.mit.edu/pub/cqncbp3l

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, *43*(S1). https://doi.org/10.1111/test.12279

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association, 79(387), 531–554. https://doi.org/10.1080/01621459.1984.10478080

Data Education in Schools. (2024). *Data Literacy Outcomes in the Scottish Curriculum for Excellence, by level and topic.* Data Education in Schools, University of Edinburgh. https://dataschools.education/resource/data-across-the-curriculum-early-years-primary-and-secondary/ (Accessed 8 December 2025)

Data Science Learning Progressions. (2025). *The K-12 Data Literacy and Data Science Learning Progressions.* https://datasciencelearning.org/ (Accessed 8 December 2025)

Davies, N., & Sheldon, N. (2021). Teaching statistics and data science in England's schools. *Teaching Statistics*, *43*(S1), S52–S70. https://doi.org/10.1111/test.12276

DfE. (2025a). 'Curriculum and Assessment Review: Building a world-class curriculum for all'. Department for Education. https://assets.publishing.service.gov.uk/media/690b96bbc22e4ed8b051854d/Curriculum_and_Assessment_Review_final_report_-_Building_a_world-class_curriculum_for_all.pdf

DfE. (2025b). 'Government response to the Curriculum and Assessment Review'. Department for Education. https://assets.publishing.service.gov.uk/media/690b2a4a14b040dfe82922ea/Government_response_to_the_Curriculum_and_Assessment_Review.pdf

DfE. (2014). Mathematics Programmes of Study: Key Stage 4. National Curriculum in England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/331882/KS4_maths_PoS_FINAL_170714.pdf.

DfE. (2013a). Mathematics Programmes of Study: Key Stages 1 and 2. National Curriculum in England. https://assets.publishing.service.gov.uk/media/5a7da548ed915d2ac884cb07/PRIMARY_national_curriculum_-_Mathematics_220714.pdf

DfE. (2013b). Mathematics Programmes of Study: Key Stage 3. National Curriculum in England. https://assets.publishing.service.gov.uk/media/5a7c1408e5274a1f5cc75a68/SECONDARY_national_curriculum_-_Mathematics.pdf

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data Moves. *Technology Innovations in Statistics Education*, 12(1). http://dx.doi.org/10.5070/T5121038001 Retrieved from https://escholarship.org/uc/item/0mg8m7g6

Ingram, J., Stiff, J., Cadwallader, S., Lee, G., & Kayton, H. (2023). PISA 2022: National report for England (1838705228). Retrieved from https://www.gov.uk/government/publications/pisa-2022-national-report-for-england

Jaques, L., & Joubert, M. (2024, November 16). *A vision for meaningful data education across the four nations of the UK*. The Joint Mathematical Council of the United Kingdom. https://www.jmc.org.uk/2024/11/16/a-vision-for-meaningful-data-education-across-the-four-nations-of-the-uk/

Jones R., & Dudzic, S. (2024). *Key Recommendations for the UK Statistics Curriculum.* The Royal Statistical Society. https://rss.org.uk/RSS/media/File-library/Policy/2024/Recommendations-for-statistics-curriculum-in-the-UK-full-paper-final.pdf

Kathotia, V., Ward-Penny, R., & Smith, C. (under review). Data Science, Mathematics and the Cohesive Potential of a Critical Questioning Paradigm

Makar, K., & Rubin, A. (2020). Informal statistical inference revisited. In Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9). International Statistical Institute. https://icots.info/icots/9/proceedings/pdfs/ICOTS9_8C1_MAKAR.pdf

Pittard, V. (2018). *The Integration of data science into the primary and secondary curriculum*. Royal Society Advisory Committee on Mathematics Education. https://royalsociety.org/-/media/policy/publications/2018/18-07-18-the-integration-of-data-science-in-the-primary-and-secondary-curriculum.pdf

Smith, C., Kathotia, V., Ward-Penny, R., Howson, O., & Wermelinger, M. (2023). *Mathematical and Data Literacy: Competencies and curriculum implications at the intersection of mathematics, data science, statistics and computing*. The Open University. https://royalsociety.org/-/media/policy/projects/maths-futures/intersection-mathematics-data-statistics-computing.pdf

Tahurangi. (2026, January). *Mathematics and Statistics Curriculum.* New Zealand Government, Ministry of Education. https://newzealandcurriculum.tahurangi.education.govt.nz/new-zealand-curriculum-online/new-zealand-curriculum/learning-areas/mathematics-and-statistics-curriculum/5637165587.c (Accessed 4 January 2026)

The Royal Society. (2023). *A new approach to mathematics and data education*. https://royalsociety.org/-/media/policy/projects/maths-futures/new-approach-to-mathematics-and-data-education.pdf?la=en-GB&hash=8FFA5FD36A6E5606C834AE38DB33B37C

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. doi:10.15353/joci.v12i3.3275

Appendix 1 Primary anchoring experiences

extracts from (ACME PEY Expert Panel, 2025) document

Enquiry- based, cross curricular maths teaching like this [as envisaged in this document] requires a major change from current primary practice, where maths is separate from other subjects, with a knowledge-based 'coverage' approach and content dominated by assessments. This approach to data education would require restructuring the primary curriculum, freeing time for enquiry-based work, with revised assessments and considerable professional development.

Examples of selected enquiries or 'Anchoring Experiences':

Phase	Subject Context	Anchor Experience	Key Learning Focus
EYFS – Y1	Citizenship / Understanding the World	Class vote or weather tally	Representing preference or change; fairness; counting as communication
Y2 – Y3	Science	Plant growth or shadow-length investigation	Measuring, recording, noticing variation
Y3 – Y4	Geography	Local litter-mapping or traffic survey	Spatial reasoning; frequency and comparison

Phase	Subject Context	Anchor Experience	Key Learning Focus
Y4 – Y5	Computing / PSHE	Digital-wellbeing log or screen-time tracker	Ethical use of data; self-monitoring; simple databases
Y5 – Y6	Science / Citizenship	Energy audit or biodiversity survey	Representing systems; multi-variable data; sustainability links

Conceptual Structure and Progression of Data Thinking

Data education needs a coherent conceptual spine — a progression of ideas about quantity, variation, and relationship that connects intuitive noticing in early years to statistical reasoning in upper primary. This structure allows data to integrate with other mathematical strands (number, measure, geometry, modelling) rather than sitting apart as a 'miscellaneous' topic.

Concept	Early manifestation	Later development
Classification & Category	Sorting and grouping tangible objects	Defining variables and categories for enquiry
Discrete & Continuous Quantity	Counting vs measuring	Understanding scale, interval, and approximation

Concept	**Early manifestation**	**Later development**
Comparison & Relationship	More than / fewer than	Difference, ratio, rate, and correlation
Variation & Distribution	Expecting small differences	Interpreting spread and range
Aggregation & Representation	Combining counts	Using averages and graphical density to model data
Uncertainty & Inference	Intuitive prediction ('maybe')	Drawing conclusions with evidence and caution

Appendix 2. ProDaBi – Project Data Science and Big Data in German schools

Project Data Science and Big Data at School (ProDaBi)[1] is a project that has been run since 2018 by computer science education and mathematics education researchers at Paderborn University. The project is funded by the foundation Deutsche Telekom Stiftung.

ProDaBi has been developing and trialling data science education materials (and some associated professional development) intended to raise students' awareness of data as a model of the world, data processing, exploration of data, and to teach them how to use data-driven machine learning (ML) for making predictions and decisions. The project started with a year-long trial with Grade 12 (Year 13) students in two schools in Paderborn, with three-hour weekly sessions, run over seven months in the school year 2018/2019 [2]. In subsequent years materials have been developed and trialled lower down the age range, first with teaching modules for lower secondary aged students, and then into the top end of primary school. Lessons have been typically taught by computer science teachers.

ProDaBi offers an example of how a computing-focus approach to data science is presently being enacted (our Model 2 for adding content to computing). It also embodies the paradigm that machine learning is a fundamental component of data science which can be communicated to school-aged students.

Predictive modelling as a fundamental idea in data science education

A starting premise for the ProDaBi project is that the industrial processes of data science are distinct from the stages of the statistical modelling cycle in many important ways. Significant amounts of data may already exist which have not necessarily been collected to a plan; the domain/business context is a crucial starting point; data preparation and cleaning are separate, time-intensive steps; and training and testing a model with data is a

---

[1] https://www.prodabi.de/
[2] https://publikationen.bibliothek.kit.edu/1000127944/97090875

distinct and important phase of the data science cycle [3]. Instead of reaching a statistically rigorous conclusion, now the goal often is to develop, validate and deploy a model which can be used for context-informed classification or prediction. This in turn raises new and important issues of social responsibility and ethics.

ProDaBi developers list three aims for their activities – demystifying the data-based machine learning (ML) model building process, investigating ML by developing their own models, and reflecting on responsible use of ML. One reason that ML is seen as integral is that data sets (and real-world contexts) typically involve multiple variables and tracking the resulting multitude of correlations by hand is simply not possible. Students could be introduced to simple classification models as an example of machine learning, for example deciding whether an image is of a cat or a dog, or determining whether a lizard has come from a natural or disturbed habitat based on attributes including its mass, length of limbs, tail length, head depth and toe-pad width [4].

Pedagogical and computational models

The ProDaBi project uses the notion of *decision trees* to develop and test its predictive algorithms, where at each stage of a multi-stage process one uses a particular attribute (and a threshold) to categorise a candidate. Decision trees serve as a transparent model for students, helpful for understanding the 'machine room' of artificial intelligence (AI), compared to, say, artificial neural networks (ANN). Decision trees provide an opportunity for teachers and students to discuss potential advantages and disadvantages of automation, and the developing role of humans in AI [5].

Students meet decision trees for the first time at the top end of primary school, where they work with physical cards that name an object and list some of its attributes. This

[3] Please see slides and video at https://iase-web.org/Webinars.php?p=230307_2000
[4] Example from GAISE II report, p. 97 https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf
[5] Slide 28 at https://iase-web.org/Webinars.php?p=230307_2000

'unplugged' approach doesn't introduce them to the mathematics or computation of data-based ML but the students get to explore the heuristics and properties of the process, recognising that different algorithms can result in different outcomes and having an opportunity to compare and discuss the accuracy and biases of their choices.

The resources provided for early secondary school students involve the students using larger data sets with semi-automatic support, specifically the computing environment Common Online Data Analysis Platform (CODAP) [6] which facilitates 'drag and drop' for articulating commands. At the top end of secondary school students can use Jupyter notebooks (cell-based environments that support creation or adaptation of Python code) or Python programming to access and work with large data sets, with Python libraries to create decision trees and related visualisations. Students are also tasked to grapple with assessing rules using misclassification rates, issues of overfitting data, false positives and false negatives, growing and pruning decision trees, and evaluating their algorithms [7].

Meaningful contexts to engage students

As well as giving students experience of machine learning processes, the ProDaBi resources demonstrate the importance of context in Data Science and are based around recognisable scenarios from the real world.

At the top end of primary school ProDaBi uses the example of choosing/recommending food items (based on nutritional information). Students are given 55 data cards and green and red paper clips to label the cards. Two example cards are given below:

---

[6] https://codap.concord.org/
[7] Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. Teaching Statistics, 43(S1), S133-S142. https://doi.org/10.1111/test.12279

Figure 2 Data cards – Nutritional facts of food items (Slide 56, at https://iase-web.org/Webinars.php?p=230307_2000)

In this activity, students get to manually create a decision tree as a rule system for classifying food items, test their tree with test data, compare different trees/approaches and explore how a machine learning method might create a decision tree.

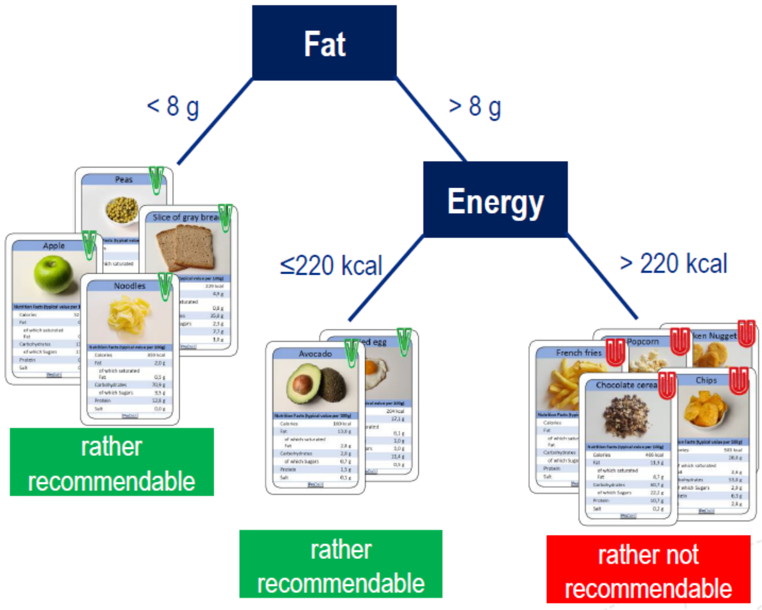

Figure 3 A decision tree for food items (Slide 64, at https://iase-web.org/Webinars.php?p=230307_2000)

One of the contexts in the materials for older students is based around social media platforms' use of online gaming behaviours to decide which advertisement is shown to a player, with data fields including the type of device someone uses to play games, the online platforms they use, the types of game, how often they plan, and so on. An authentic data set to support this resource was created by the ProDaBi developers working with secondary schoolchildren.

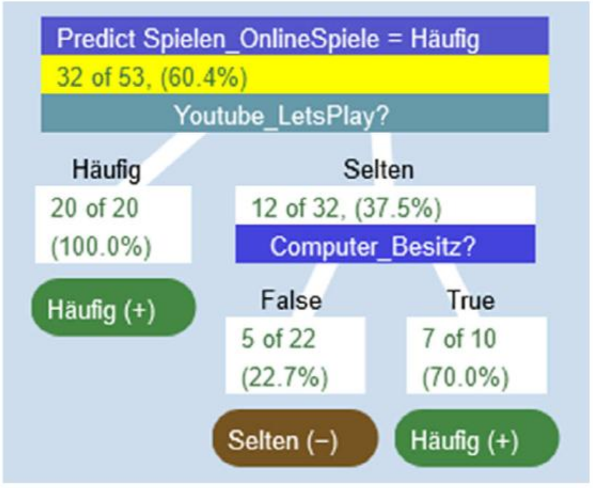

Figure 4 A student's solution creating a decision tree with CODAP (Biehler & Fleischer, Teaching Statistics. 2021;43:S140 ***Error! Bookmark not defined.****)*

Students can use the Jupyter Notebooks to focus on one menu-based action at a time or move on to use more versatile code-based actions; they can also use the notebooks to add explanations and narrative.

Benefits

- The unplugged/embodied, menu-driven and Python code-based implementations allow different age groups to explore ML.
- Relevant and authentic contexts support student engagement and sense-making.
- Students get to appreciate the role and need for computational tools, automation and ML.
- Students see and grapple with the ethical implications of ML.

Barriers

- The mathematical demand of some of the constructs (confusion matrices / two-way tables) relating to the decision trees used in some of the resources can be problematic – students need to be able to use these.
- Teaching was done by computing teachers and there are related teacher supply issues.
- Some teachers may have difficulties filling in student gaps in mathematical knowledge.

Adaptations

- Team teaching could be considered to draw in complementary knowledge across disciplines.
- The developers are experimenting with different sized data sets to draw out different aspects of ML.

The Open University