

RECOMMENDATIONS
FOR EFFECTIVE
STATISTICS
ASSESSMENT:
**A RAPID REVIEW
OF EVIDENCE**
OCTOBER 2026

Recommendations for Effective Statistics Assessment: A rapid review of evidence

Darren Macey & Lucy Rycroft-Smith
October 2026

*Produced by the **Centre for Statistics and Data Science Education**
on behalf of the **Joint Mathematical Council of the United Kingdom**
and the **Royal Statistical Society***

Acknowledgements:

We would like to thank the Royal Statistical Society and the Joint Mathematical Council of the UK for commissioning this report. We are grateful to the members of the Education Policy and Public Engagement Group of the RSS and the trustees of the JMC, along with the directors of the Centre for Statistics and Data Science Education, for reviewing the draft report.

This rapid review was funded in part by the Royal Society.

FUNDED BY

THE
ROYAL
SOCIETY

Contents

Executive summary.....	1
Introduction	5
Section 1: Key organising concepts from statistics education research	6
1.1 Recommendations from previous reports	10
Section 2: Recommendations for assessment design.....	13
2.1 Literacy: developing statistical thinking and data literacy	13
2.2 Technology: software tools and applications	16
2.3 Enquiry: using the statistical problem-solving cycle	18
2.4 Authenticity: Using real-data and relevant contexts.....	19
2.5 Responsibility: ethics and data stewardship	23
Section 3: Review of current approaches to assessment in the four nations of the UK	24
3.1 Background	24
3.2 Review and reform.....	24
3.3 Review of UK statistics assessment against the five features	25
3.4 Summary of assessment.....	28
Section 4: Some examples of statistics assessment in other jurisdictions	30
4.1 Overview	30
4.2 Aotearoa New Zealand	30
4.3 The USA: Levels of Conceptual Understanding in Statistics Project (LOCUS) ...	31
Conclusion	33
References	35
Appendix: Approach to assessment review	40

Executive summary

This report provides evidence on effective statistics assessment at school level, and ways in which current assessment practice across the four nations of the UK is aligned or not to such evidence. It draws on research in statistics education, recent reports from the Royal Society, the Royal Statistical Society and the Joint Mathematical Council of the UK, international practice, and a rapid review of current mathematics and statistics qualifications in the four nations of the UK.

Organising concepts

Statistics education research has long distinguished **statistical literacy** (using the language and tools of statistics and critically evaluating statistical claims), **statistical reasoning** (making sense of and arguing with statistical information) and **statistical thinking** (understanding the whole investigative process, including handling variability and uncertainty). More recently, **data literacy** has emerged alongside these: solving real problems with real data, including data cleaning, technology use and ethics. We consider these to be organising concepts for this report. To visualise these useful but often overlapping and contested ideas from the literature, we propose a 'twin peaks' model in which statistical literacy underpins both statistical thinking and data literacy. The intention is that students of all ages and attainment levels revisit and climb both peaks iteratively, rather than younger or lower-attaining students remaining at the lower layer. We recognise also that these concepts might be hard to imagine in practice, and so we have drawn out five particular features from the literature that illustrate or exemplify the move towards coherently teaching and assessing statistical literacy, reasoning and thinking, and data literacy.

The five features

From this evidence base, the report synthesises and proposes five features of evidence-informed statistics education that should apply to statistics assessment:

1. **Literacy:** Focus on the (explicitly named and clearly defined) concepts of statistical literacy, statistical reasoning and/or statistical thinking, and data literacy (this is the overarching principle, and includes all of the below)
2. **Technology:** Embed the use of software tools to support enquiry and modelling; use the same tools in teaching and assessment

3. **Enquiry:** Focus on statistical problem-solving using the PPDAC (problem, plan, data, analysis, conclusion) cycle or equivalent
4. **Authenticity:** Use real data to solve realistic¹ statistical problems wherever possible, explicitly including a diverse range of cultural contexts including those relevant to local and global community; reduce the emphasis on technical fluency and by-hand calculations
5. **Responsibility:** Explicitly include ideas of ethics and data stewardship (using data carefully and responsibly to minimise risk and harm)

Current practice in the UK

The review found that no nation's assessment is currently well aligned with the five features, although there are pockets of good practice. Across the four nations:

- Paper-based exams dominate and are built mainly from short and medium structured items with point-based analytic mark schemes. These reward multi-step mathematical working, such as calculation-heavy significance tests, while sense-making typically earns only the final one or two predictable marks. No reviewed paper allocated substantial marks to extended statistical argument or demonstration of statistical reasoning.
- Specifications do sometimes refer to statistical thinking and data literacy, but this is rarely reflected in exam items, even in specialist statistics qualifications.
- Access to technology is generally limited to calculators. The main exception is Scotland, where some components of assessment require or allow statistical software.
- Real data and cross-curricular contexts appear sometimes, but scaffolding means students are rarely given the opportunity to solve realistic statistical problems. Pre-release materials may create opportunities for authenticity, but the item formats limit these opportunities.
- There is almost no assessment of ethics or data stewardship.
- Scotland's use of projects and internally assessed tasks in some qualifications offers more scope across several features. However, this is not universal, its written papers still favour mathematical working, and the quality of teacher-selected tasks could not be judged in this review.

¹incorporating real or realistic data and attempting to mimic aspects of the work of professional or academic statisticians.

International examples

Aotearoa New Zealand's NCEA includes internally assessed statistical enquiries in which students choose their own investigative question, use software, and are assessed with levels-of-response rubrics that give explicit weight to important concepts such as variation. The US LOCUS project shows that paper-based items, including well-designed multiple-choice questions, can assess conceptual understanding across the statistical problem-solving cycle with minimal calculation. Its items are balanced across formulating questions, collecting data, analysing data and interpreting results. Neither model meets all five features, but both offer aspects and approaches the UK nations could adapt.

Key recommendations

The report makes ten recommendations:

- 1. A range of validated and reliable item formats and contexts should be used to assess statistical thinking, statistical reasoning, statistical literacy, and data literacy.**
- 2. Assessment items intended to assess statistical thinking should require and isolate it.**
- 3. Assessment of statistical thinking, reasoning and literacy should be resourced to reflect its true cost.**
- 4. The technology used in statistics assessment should be aligned with the technology used in statistics teaching.**
- 5. Pilot and evaluate hybrid models of technology-supported assessment.**
- 6. A research agenda should be prioritised that establishes design guidelines for good practice in technology-supported summative statistics assessment.**
- 7. Statistics assessment should include opportunities for students to apply the complete statistical problem-solving cycle to authentic statistical problems and/or investigations.**
- 8. Statistics assessments should make use of real data and contain authentic tasks that mirror the work of practising statisticians, supported by technology tools where possible.**
- 9. Statistics assessment design should foreground conceptual work and reduce emphasis on procedures and techniques.**
- 10. Data stewardship and ethical use of data should be embedded within statistics assessments.**

Priorities for action

Three structural barriers stand in the way of better alignment with the five features; addressing them is likely to be the essential first step. They are:

1. Access to statistical software within assessments.
2. Paper-based exams limit opportunities for students to conduct investigative work and engage with the full PPDAC cycle.
3. Short-structured items and point-based analytic mark schemes reward **mathematical working** at the expense of **statistical thinking**.

Introduction

This report considers the central question of what constitutes effective statistics assessment, supported by the latest high-quality research, evidence of practice from across the UK and other jurisdictions, and expert synthesis of knowledge, all applied in the context of the constraints of current educational policy and practice. In particular we emphasise that while research is important, it is not the only type of evidence that is important to the questions being considered here; we have also weighed other types of evidence and taken seriously a range of diverse types of information in understanding best practice in statistics assessment at school level.

This report begins by considering some key organising concepts in statistics education across school education that are recommended as foundational for assessment practices. While the focus is on summative assessment (evaluative; an assessment of learning) here, it is recognised that the boundary between summative and formative assessment (diagnostic, assessment for learning) is not distinct, since assessment items/artefacts are often used for multiple purposes, students receive feedback in different ways than intended, and accountability incentives in the education system can merge the two.

Another false binary relevant here is curriculum and assessment, which are interrelated in significant and complex ways. Consequentially, we reference curriculum throughout this report, considering in particular ways that evidence might impact on the interplay between curriculum and assessment. Related to this, the report assumes that statistics remains embedded within the broader mathematics curriculum, in line with the Royal Society's *A new approach to mathematical and data education*. The scope of the report therefore is limited to mathematics qualifications, and the small number of statistics qualifications available to study in the four nations of the UK. We recognise that there are strong arguments that statistics can be (and often is) assessed through other subjects with application in authentic contexts. We also acknowledge that good practices in statistics assessment may well exist across the UK currently in other subjects; however, this was beyond the scope of this report.

Section 1: Key organising concepts from statistics education research

In the context of statistics education we are challenged to ‘teach and assess what we claim to be important’ (delMas, 2002, p. 2) and ‘assess what you value’ (Chance, 2002, p. 9). The second Guidelines for Assessment and Instruction in Statistics Education (GAISE) report released in 2020 suggests that statistics education should focus on:

- 1. The importance of asking questions throughout the statistical problem-solving process (formulating a statistical investigative question, collecting or considering data, analysing data, and interpreting results), and how this process remains at the forefront of statistical reasoning for all studies involving data*
- 2. The consideration of different data and variable types, the importance of carefully planning how to collect data or how to consider data to help answer statistical investigative questions, and the process of collecting, cleaning, interrogating, and analysing the data*
- 3. The inclusion of multivariate thinking throughout all Pre-K–12 educational levels*
- 4. The role of probabilistic thinking in quantifying randomness throughout all levels*
- 5. The recognition that modern statistical practice is intertwined with technology, and the importance of incorporating technology as feasible*
- 6. The enhanced importance of clearly and accurately communicating statistical information (Bargagliotti et al., 2020, p. 2)*

These goals of statistics education from the last few decades have been summarised as a focus on **statistical literacy**, **statistical reasoning**, and **statistical thinking** (e.g. Ben-Zvi & Garfield, 2004; Garfield & Franklin, 2011; Saidi & Siew, 2022), deliberately framed in contrast to solely teaching ‘skills, procedures, and computations, which do not lead students to reason or think statistically’ (Ben-Zvi & Garfield, 2004, p. 3). These three ideas are closely related, but distinct, and too often

conflated or unclear (Chance, 2002). According to Garfield and Franklin (2011), the following definitions apply:

- **Statistical literacy** is understanding and using the basic language and tools of statistics: words, symbols and representations of data, as well as interpreting and critically evaluating statistical products (for example graphs and charts in the media) (Gould, 2017), and is foundational to statistical reasoning and statistical thinking, as well as data literacy.
- **Statistical reasoning** is connecting one statistical concept to another, making arguments with statistical ideas and making sense of statistical information, understanding processes and interpreting results, and includes both informal and formal reasoning
- **Statistical thinking** is understanding the entire investigative process, including question posing, collecting data, modelling, drawing inferences, testing assumptions, recognising and explaining variability, handling uncertainty, paying attention to data provenance, making justified conclusions, and understanding and using context. Chance (2002, p. 4) adds moving beyond what is taught to ‘spontaneously question and investigate’.

Relatedly, **data literacy** has been described as solving realistic problems with data; this includes asking and answering real-world questions from large and small data sets through an inquiry process, thinking about the ethical use of data, and selecting, cleaning, analysing, visualising, critiquing and interpreting data as part of communicating stories from data and also as part of a design process (Wolff et al., 2016).

The difference between data literacy and statistical literacy is unclear in current research; Wolff suggests that statistical literacy is part of data literacy. Data literacy has also been called **data science literacy** (e.g. Dichev & Dicheva, 2017), since for many researchers it includes a focus on **data science**, defined as an interdisciplinary field that synthesises and builds on statistics, informatics, computing, communication, management, and sociology to study data and its environments (Cao, 2017). In recent research there has been a shift towards making connections between statistics and data science, and potentially recontextualising some elements of the statistics curriculum to emphasise the ways in which statistical literacy ideas underpin data science. For example, the American Statistical Association’s Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) A Framework for Statistics and Data Science Education

(Bargagliotti et al., 2020) suggests an increased emphasis on the use of technology, multivariate data, and skills such as data cleaning.

This suggests the idea of twin pathways or peaks of statistical learning and teaching. On the left hand side of Figure 1, this is shown as beginning with the foundational work of statistical literacy (recognising and exploring data) and then moving through to statistical reasoning (making sense of that data) and statistical thinking (using data in more sophisticated ways; investigating, contextualising and handling variability in data). The boundaries between these elements are not hard; they are porous and overlapping (delMas, 2002). On the right-hand side, with some overlap, students also begin with the foundational work of statistical literacy (recognising and exploring data) but this leads in a slightly different direction, this time to data literacy (inquiry using real datasets, developing technical skills, and considering ethics). A simplified way of looking at these concepts could be as mathematics and ICT silos in the enacted curriculum, i.e. as work in the mathematics classroom on the left, and in the ICT classroom on the right. This model emphasises that statistical literacy is foundational to both, and there is likely much overlap.

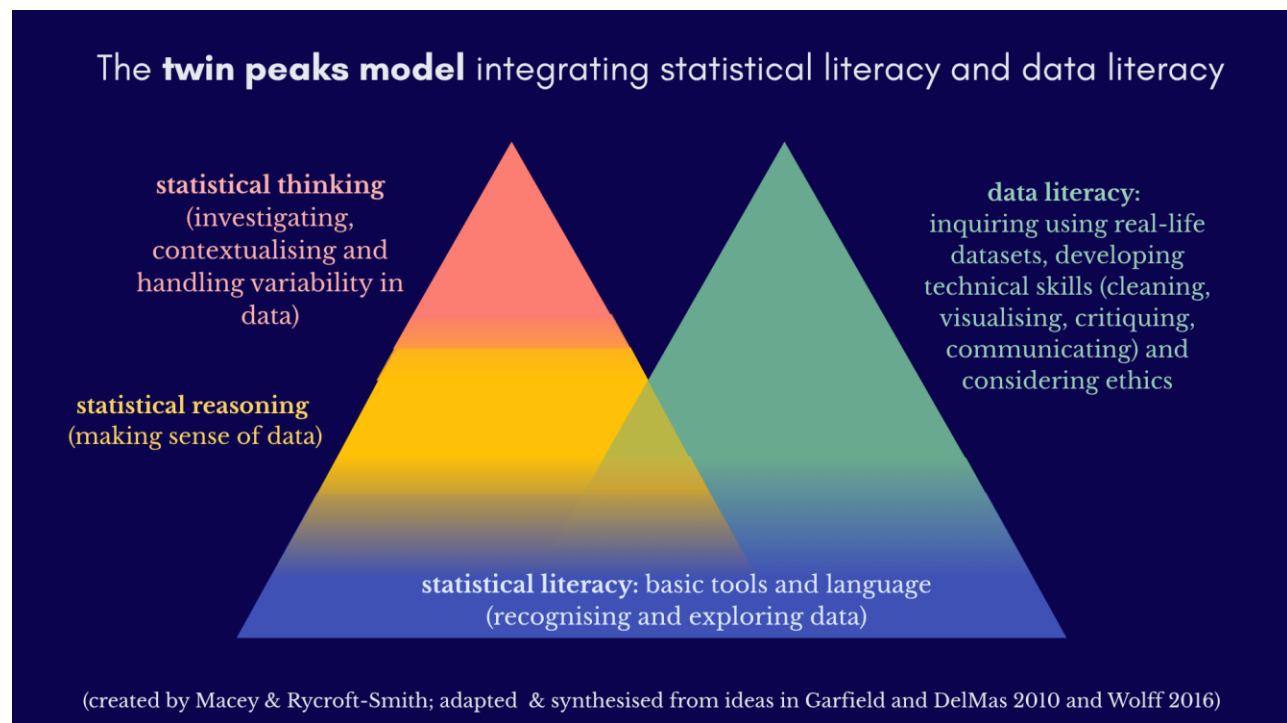


Figure 1: The twin peaks model integrating statistical literacy and data literacy

This hierarchy could also be misinterpreted as young students (or lower attaining students) working at the 'bottom' of the peak, and older students (or higher attaining students) at the top. Instead, the evidence suggests an **iterative ascent model**, where students of all ages and attainment levels pursue pathways towards more and more sophisticated statistical work as they add new skills and knowledge, looping back around as needed to grow their understanding of statistics and data as an approach (e.g. delMas, 2002).

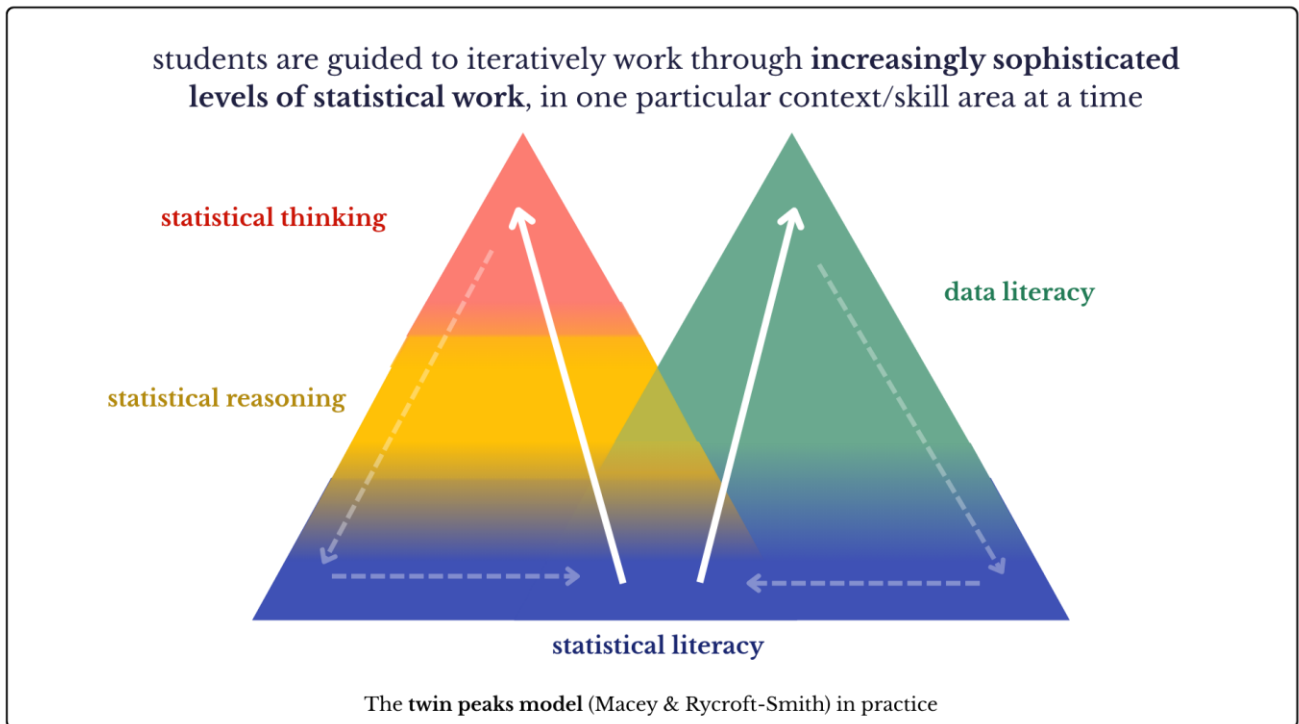


Figure 2: The twin peaks model in practice, showing an iterative ascent cycle

Other researchers have argued that these conceptions relate in different ways; for example Gould (2017) proposes that the proliferation of data in our daily lives has merged ideas of statistical literacy and data literacy such that they could be considered fully integrated.

Regardless, the goals of developing data literacy and statistical thinking in statistics education, supported by evidence, are constituted by:

- Understanding and using the basic language and tools of statistics
- Becoming a critical consumer of statistical products
- Reasoning with statistical ideas and making sense of statistical information
- Examining data provenance and trying to explain variability

- Engaging in the full inquiry process wherever possible, including formulating questions, collecting/cleaning data, analysing, and drawing conclusions
- Clearly communicating statistical information and/or conclusions drawn from data
- Grounding all of the above in realistic problem contexts
- Developing curiosity and independence as a statistician

In this section we have looked at the organising concepts of statistical thinking, statistical reasoning, statistical thinking, and data literacy, finding them to be overlapping and often difficult to understand in practice. In the next section, therefore, we consider recommendations from recent reports in statistics education that have illuminated elements of these concepts and finish with a list of features that we consider to be well supported by evidence as foundational to effective statistics education practice. Because the evidence also points to a strong need for coherence between curriculum and assessment, we consider evidence more broadly around statistics education and not just confined to assessment.

1.1 Recommendations from previous reports

The Royal Statistical Society, The Royal Society, and The Joint Mathematical Council of the UK have each recently published reports that set out similar perspectives on statistics education, with recommendations for curriculum and assessment, which are shown in Table 1.

Table 1: Recommendations for statistics education (curriculum and assessment), showing similarities and differences

Institution	Joint Mathematical Council for the Royal Society	Royal Society	Royal Statistical Society
Report	<i>Data education and mathematics 5–16</i> (Kathotia & Smith, 2026)	<i>A new approach to mathematical and data education</i> (The Royal Society, 2024)	<i>Key recommendations for the statistics curriculum in the UK</i> (Jones & Dudzic, 2024)
Recommendations in common	Overarching: focus on the (explicitly named) concepts of statistical thinking and data literacy; this report also references data education (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)	Overarching: focus on the (explicitly named) concept of data literacy and statistical thinking; this report also references the broader concept of quantitative literacy (application of mathematical and data skills to common, real-world, quantitative problems in a range of educational, employment and everyday contexts)	Overarching: focus on the (explicitly named) concepts of statistical literacy and statistical reasoning
	Embed the use of software tools, integrating technology into students' experience of data	Embed the use of interactive software tools	Embed the use of software tools, especially for visualisation
	Focus on statistical problem-solving using the full statistical investigative cycle / PPDAC cycle	Focus on statistical problem-solving using the problem-solving cycle/ PPDAC cycle	Focus on statistical problem-solving using the investigative cycle/PPDAC cycle
	Emphasise statistical sense-making and storytelling with data, as opposed to practising techniques and avoid de-contextualising statistical work	Emphasise contextualised sense-making as opposed to technical fluency	Emphasise reasoning and sense-making as opposed to practising techniques
	Emphasise connections with other subjects	Emphasise connections with other subjects through developing quantitative literacy	Emphasise connections with other subjects
	Use real data to solve authentic statistical problems wherever possible	Apply statistical principles in meaningful, real-life contexts, using real data to solve authentic statistical problems wherever possible	Apply statistical principles in meaningful real-life contexts, including a range of cultural contexts

Distinct recommendations	When collecting data, begin with a purpose relevant to students' locality and context.	Develop probabilistic reasoning	Reduce emphasis on mathematical / by-hand calculations to instead focus on concepts and contexts
	Explicitly include ethics and data stewardship (using data carefully and responsibly to minimise risk and harm)	Offer quantitative literacy problems that don't have a single correct answer	
	Draw on elements of mathematics, computing and statistics/data toolkits		
	Emphasise the creative aspects of statistics including storytelling and design		

From analysis of these sources and the wider evidence base, we therefore synthesise five features of evidence-informed statistics education that should be evident in statistics assessment:

1. Literacy: Focus on the (explicitly named and clearly defined) concepts of statistical literacy, statistical reasoning and/or statistical thinking, and data literacy (this is the overarching principle, and includes all of the below)

2. Technology: Embed the use of software tools to support enquiry and modelling, and use the same tools in teaching and assessment

3. Enquiry: Focus on statistical problem-solving using the PPDAC (problem, plan, data, analysis, conclusion) cycle or equivalent

4. Authenticity: Use real data to solve realistic statistical problems wherever possible, explicitly including a diverse range of cultural contexts including those relevant to local and global community; reduce the emphasis on technical fluency and by-hand calculations.

5. Responsibility: Explicitly include ideas of ethics and data stewardship (using data carefully and responsibly to minimise risk and harm)

We explore the recommendations and expressions of these features below and return to them in sections 3 and 4 examining statistics assessment in the four nations of the UK, as well as current practice in some jurisdictions exemplifying good practice.

Section 2: Recommendations for assessment design

In this section we consider evidence from research for the design of assessment in statistics education in more detail by looking at each of the five features in turn, synthesising the research into recommendations.

2.1 Literacy: developing statistical thinking and data literacy

Variety in assessment items is important for developing statistical reasoning and statistical thinking. MacGillivray (2010) argues that in both formative and summative assessment, a greater range of contexts and assessment styles provide more opportunities for students to engage reasoning in different ways balancing the procedural, creative, and interpretative elements of the discipline of data analysis. They make the case that no single style of item is appropriate to assess all aspects of the statistics curriculum and so a range of assessment and item formats are necessary so that each component of the statistics curriculum's aims is assessed using valid methods. Gal and Garfield (1997) made a similar argument, claiming that multiple approaches to assessment are needed to assess specific skills, understanding of concepts, and quality of thinking, reasoning, and communication.

In a systematic review of assessment instruments in research studies into both students' and teachers' data literacy, Cui et al. (2023) identified a number of formats used including: multiple-choice, constructed response, digital game-based, interview, observation framework, and student-generated posters. While these approaches were used to measure the impact of data literacy interventions, only two of the studies reported explicitly on measures of validity and reliability of the assessment and both of these related to multiple-choice assessment. The researchers suggest that assessment of data literacy is an emerging field and there remains a gap in research related to development of high-quality assessments in this area.

Garfield and Franklin (2011) make a series of recommendations related to item format, suggesting:

- Multiple-choice items should have meaningful distractors that reveal common errors in learning, reasoning or conceptual understanding
- Decisions about whether to use multiple-choice or free-response items need to be informed by the purpose of the assessment

- Free-response items should be used to allow students to create, explain, and communicate their understanding.

Isolating statistical thinking as a construct within mathematics

In the previous section we considered statistical thinking as a valuable and distinct idea in mathematics education, which draws on investigating, contextualising and handling variability. In order to assess students' statistical thinking, items must be designed so that students cannot answer them by relying purely on mathematical reasoning through the application of processes and procedures (Bargagliotti & Groth, 2016), even if those procedures engage valuable components of mathematical reasoning.

Example 1:

The ratio of the mean of list P: the mean of list Q is 4:7

List P: 12, 28, 16, 25, 39

List Q: 39, 24, 49, 55, x, 36, 43,

What is the value of x?

In this example, the mean algorithm is used as a basis for a mathematical problem and would meet the requirements of GCSE mathematics in England across two dimensions. In terms of curriculum content, it assesses knowledge of how to calculate the mean. By combining the mean algorithm, ratio notation, and symbolic algebra it can also be allocated to the problem-solving assessment objective. While an interesting and potentially challenging mathematical task, it does not require any statistical thinking to solve.

Example 2:

Problem 1: Assume a die is 'fair.'

Question: If a die is rolled 10 times, how many times will we see an even number on the top face?

Problem 2: You pick up a die.

Question: Is this a fair die? That is, does each face have an equal chance of appearing? (Bargagliotti et al., 2020, p. 10)

This issue can also arise through the way in which assessment items are framed. In the second example taken from GAISE II, Problem 1 is considered a mathematical question in which students apply a mathematical rule without considering broader concepts of probability. By contrast, Problem 2 is a statistical question that still requires students to use the same mathematical probability model but to apply this

to the design of a simple probability experiment. The authors suggest that statistics assessment should be weighted towards the latter, claiming that probability is the foundation for predictions in statistics, and therefore statistical uses of probability should be assessed. Garfield and Franklin (2011) suggest that statistics assessment should require students to recognise and understand the concept of variability, while maintaining a balance between assessing items assessing students' understanding of probability concepts and their understanding of statistics concepts.

Davies and Marriot further illustrate a key tension between mathematical and statistical questions, claiming that 'Mathematical statistics questions often take a long time to set but are fairly quick and objective to mark. Statistics questions in general can be lengthy to set and lengthy to mark, and need careful marking schemes.' (Davies & Marriott, 2010, p. 13). This is particularly important to consider in the context of assessment systems with constrained budgets, or that operate in a commercial environment where the differing costs associated with developing and administering effective assessment of statistical thinking compared to mathematical questions may incentivise more of the latter.

Bargagliotti and Groth (2016) also explore this tension, demonstrating that even when items are designed to assess statistical thinking, subtle choices may mean that it is possible to construct a viable answer that only draws on mathematical reasoning. They highlight that care must be taken to ensure that quantities and contexts are chosen such that viable mathematical approaches are mitigated.

Example 3:

An organisation runs a competition to guess the number of red marbles in a jar. Competitors are allowed to take a random sample of 16 marbles from the jar to help them with their guess.

The true proportion of red marbles in the jar is 33.6%.

Should you be concerned if none of the samples contain exactly 33.6% red marbles?

(Adapted from Bargagliotti & Groth, 2016)

In this example, the anticipated **statistical** answer refers to variability between samples, but a valid **mathematical** answer given unexpectedly by a participant in the study referred to the mathematical impossibility for any sample of 16 to have 33.6% red marbles as this value would fall between 5 and 6 red marbles. The researchers suggest that assessment items should be constantly evaluated and refined based on analysis of student responses to ensure that they are assessing the intended statistical concepts. To support this process, assessment designers should

state the conceptual focus of each item in associated documentation (Gibeau & Cousineau, 2025), for example in marking rubrics.

Recommendation 1: A range of validated and reliable item formats and contexts should be used to assess statistical thinking, statistical reasoning, statistical literacy, and data literacy. For example, by combining formats that together capture procedural skills, understanding of concepts, and the quality of students' reasoning, interpretation and communication.

Recommendation 2: Assessment items intended to assess statistical thinking should require and isolate it. Items should be designed so that students cannot reach a fully credited answer through mathematical procedures alone. Every such item should require students to engage with at least one distinctly statistical element, for example: interpreting variability, reasoning about data in context, or making data-informed judgements.

Recommendation 3: Assessment of statistical thinking, reasoning and literacy should be resourced to reflect its true cost. Statistical questions may typically take longer to set and to mark than mathematical ones, and require carefully constructed mark schemes.

2.2 Technology: software tools and applications

Many researchers have pointed out that high-stakes assessment exerts considerable influence on classroom practice (Ben-Zvi, Gravemeijer & Ainley, 2018; Burkhardt & Schoenfeld, 2018; Garfield & Franklin, 2011; Macey & Rycroft-Smith, 2022), which suggests that **coherence** (alignment between high-stakes assessment and the intended curriculum) is an important consideration. One possible approach to supporting this coherence is to take a **learning environment perspective** on design, where students have access to the same tools and ways of working in both classroom and assessment, which we noted earlier in the section on technology. For example, encouraging the use of real data or technology in classrooms is unlikely to have impact if these are not also used in assessment. This was demonstrated in the inclusion of a large data set in England's Mathematics A-level, which has been reported as being unsuccessful in large part because the computer tools needed to work with the data were not available in the assessment (e.g. Davies & Sheldon, 2021; Ridgway & Nicholson, 2019).

Chan et al. (2015) highlight in their systematic review a need for more research into technology supported assessment of statistical reasoning. They suggest that technology use can support greater understanding of statistical concepts, and more extensive engagement in the full statistical problem-solving process while reducing the burden of calculation and graph construction in favour of exploring, analysing, and interpreting data. The use of technology in the form of software tools is explored throughout the *International Handbook of Research in Statistics Education* (Ben-Zvi, Makar & Garfield, 2018) and supports the findings of Chan et al., while also acknowledging that issues with technology use remain. These include:

- Equity of access and infrastructure limitations
- Teachers' knowledge and skills related to both software use and effective pedagogy
- Rapidly evolving software and hardware
- Lack of alignment between curriculum and assessment in relation to technology use
- Limited space in the curriculum and limited availability of classroom time.

A strong theme within the handbook is that technology use on its own is insufficient, and requires well-designed tasks and activities delivered by confident teachers with sufficient pedagogical knowledge to use the tools effectively.

Some common statistical tools used with students in the research literature include: Tinkerplots; CODAP; and iNZight. These all have the advantage of being designed with statistics classrooms in mind. Other tools used by professional statisticians such as R Studio and Python have also been used in research, particularly with older students.

GAISE II argues explicitly that assessments should align the technology used for statistical computation with the technology used for teaching. Mismatches such as having computer tools available in the classroom but only calculators in the assessment should be avoided (Bargagliotti et al., 2020).

This supports the idea that technology tools should be embedded in assessment. However, it does not automatically imply a requirement for fully computer-based testing. An example of a hybrid approach can be found in a small research study conducted by Chan and Ismail (2014) who designed and validated hybrid assessment items in which students answered traditional paper-based items alongside related items that utilised Geogebra spreadsheets and worksheets.

Recommendation 4: The technology used in statistics assessment should be aligned with the technology used in statistics teaching. Students should have access to the same statistical software in summative assessment as in the classroom, and mismatches such as software in lessons but calculators only in exams should be avoided.

Recommendation 5: Pilot and evaluate hybrid models of technology-supported assessment. Awarding bodies should trial hybrid formats that combine paper-based items with linked software-based tasks and evaluate their impact, focusing on developing mark schemes that credit students' interpretation of results and their choices about how to use the tools, rather than the outputs the software produces.

Recommendation 6: A research agenda should be prioritised that establishes design guidelines for good practice in technology-supported summative statistics assessment. For example, including trialling hybrid models that combine paper-based and software-based tasks along with a consideration of the limitations of existing school infrastructure.

2.3 Enquiry: using the statistical problem-solving cycle

The importance of the statistical problem-solving cycle within statistics education is well-established both within the UK (e.g. Porkess, 2013) and in the wider international literature (e.g. Ben-Zvi, Makar & Garfield, 2018). Researchers suggest that the process of statistical exploration is as important as the product (Chance, 2002). GAISE II recommends that students have frequent opportunities to conduct thorough and realistic statistical investigations (Bargagliotti et al., 2020). This implies a need for some component of assessment to include full-cycle investigations in order to align assessment and intended classroom practice. This is further supported by evidence referred to in section 2.2. Garfield and Franklin (2011) recommend that statistics assessment should include opportunities to select methods for graphing and analysing data and provide interpretations of data analysis, as well as justifications for their analyses and conclusions, and opportunities to plan, conduct, and describe a statistical investigation.

Recommendation 7: Statistics assessment should include opportunities for students to apply the complete statistical problem-solving cycle to authentic statistical problems and/or investigations.

2.4 Authenticity: Using real data and relevant contexts

Bidgood (2010) argues in the context of university students that assessing statistical concepts should be based in **realistic problems**, incorporating real data and attempting to mimic aspects of the work of professional or academic statisticians. They draw attention to the opportunities for more authentic statistical activity where technology tools are part of assessment and highlight the impact on student motivation of real-world data and problems. Although aimed at university students taking introductory statistics courses, this aligns with school-level recommendations such as those in the GAISE reports discussed in previous sections. The key role within statistics education of using authentic data to solve real or realistic statistical problems is well-established both within the UK (e.g. Porkess, 2013) and in the wider international literature, stretching back many decades (e.g. Ben-Zvi, Makar & Garfield, 2018; Cobb, 1992). Garfield and Franklin's (2011) recommendations for statistics assessment support the call for authenticity in assessment. They also suggest that assessment items asking for computations (e.g. calculating the mean) should have a context demanding more statistical thinking than just computed answers, and any computed answer should have a meaningful purpose.

One of the key questions that statistics education research poses when it comes to assessment is how best to balance the assessment of skills, procedures and computations against the impetus from these organising principles to contextualise, authenticise, and broaden assessment in order to test more conceptual understanding. To help resolve this tension, Gal and Garfield present mastering procedural skills as one goal amongst eight in total, as shown in Table 2, with the remaining seven goals more conceptually oriented.

Table 2: Goals for curriculum and assessment (adapted from Gal and Garfield, 1997)

Goal 1	Understand the purpose and logic of statistical investigations
Goal 2	Understand the process of statistical investigations
Goal 3	Master procedural skills
Goal 4	Understand formally or informally the main mathematical ideas underpinning statistics
Goal 5	Understand probability and chance
Goal 6	Develop interpretative skills and statistical literacy
Goal 7	Develop an ability to communicate statistically
Goal 8	Develop useful statistical dispositions such as an appreciation of the role of randomness and chance in the real world, the value of statistical methods as scientific tools, and a critical perspective on data-based claims

GAISE II notes that ‘Regardless of the type of assessment used, assessment items should measure conceptual understanding. They must require students to use statistical reasoning with context and variability at all stages of the statistical problem-solving process.’ (Bargagliotti et al., 2020, p. 11). Garfield and Franklin (2011) recommend that statistics assessment tasks should be situated in a context for which there is a good explanation of what question is being asked and why data were collected. In order to support this recommendation, several research studies use the Statistical Reasoning Test Survey (STRS) instrument to assess students’ statistical reasoning (e.g. Saidi & Siew, 2019, 2022), which was found to be an extremely reliable instrument for assessing students’ statistical reasoning. The instrument uses constructed response items, with each asking students to ‘explain why’. Some items do test procedural knowledge, but many ask students to make a choice or engage in critique (e.g. can you organise the data in the table in different ways? Which graph represents the data better? Which average would be most appropriate for these data sets?). The marking rubric is an adaption of the SOLO (Structure of Observed Learning Outcomes) taxonomy (Biggs & Collis, 1982), which is a general hierarchy of reasoning.

Table 3: Overview of SOLO and STRS

SOLO taxonomy level (general reasoning)	STRS instrument (statistical reasoning)
Prestructural	Sub process
Unistructural	Idiosyncratic
Multistructural	Transitional
Relational	Quantitative
Extended abstract	Analytical

In Saidi and Siew's study, each answer was identified as Idiosyncratic, Transitional, Quantitative, or Analytical and given a score of 1–4 respectively. Each item was assigned to a data literacy construct, with exemplification of responses at each level. Table 4 illustrates the researcher's framework and is presented as an exemplar that could be adapted in the context of UK assessment. An alternative example can be seen in Bargagliotti et al. (2020).

Table 4. Initial Framework of the Statistical Reasoning Test Survey (adapted from Saidi and Siew, 2019, p. 539)

Construct	Sub process	Idiosyncratic	Transitional	Quantitative	Analytical
Describing data	<i>Showing awareness of display features.</i>	Shows no awareness to the displayed features (e.g. title and axis labels)	Shows only a single awareness to the displayed features	Shows some awareness to the displayed features	Shows complete awareness to the displayed features
	<i>Identifying units of data values</i>	Unable to identify units of data values	Identifies only a single unit of data values	Identifies some units of data values	Identifies units of general data values completely
Organising data	<i>Grouping or organising data</i>	Unable to group or organise data into classes	Group or organise the data into classes that are not consistent (have some flaws)	Group or organise data into classes without flaws	Group or organise data into classes without flaws and can group the data in more than one way
	<i>Summarising data in terms of centre</i>	Unable to summarise the data in terms of	Have an idea about measures of central	Summarises the data using a measure of	Summarises the data using valid and correct

		measures of central tendency	tendency but unable to summarise using a valid measure	central tendency but may have some flaws in the procedure	measures of central tendency
	<i>Summarising data in terms of measures of spread</i>	Unable to summarise the data in terms of measures of spread	Have an idea about measures of spread but unable to summarise using a valid measure	Summarises the data using a measure of spread but may have some flaws in the procedure	Summarises the data using valid and correct measures of spread
Representing data	<i>Constructing a data display for a given data set</i>	Unable to construct a data display for a given data set	Constructs a data display that is partially complete for a given data set	Constructs a data display for a given data set completely, but the display may have a few minor flaws	Constructs a data display for a given data set completely, with no flaws
	<i>Evaluating the effectiveness of data displays in representing data</i>	Compare the effectiveness of two different data displays for the same data set by using irrelevant features or reasons	Compare the effectiveness of two different data displays for the same data set using a single relevant feature	Compare the effectiveness of two different data displays for the same data set using more than one relevant feature	Comprehensively compare the effectiveness of two different data displays for the same data set
Analysing and interpreting data	<i>Reading between data</i>	Provides no or incorrect comparisons within and between data display or set	Provides only a single comparison within and between data display or set correctly	Provides local comparisons (e.g. shape/ centre/ spread) within and between data display or set correctly	Provides global comparisons (e.g. shape, centre, and spread) within and between data display or set correctly
	<i>Reading beyond data</i>	Provides inferences that are not based on the data or based on irrelevant issues	Provides inferences that are partially based on the data. Some inferences may be only partially reasonable	Provides inferences primarily based on the data. Some inferences may be only partially reasonable	Provides reasonable inferences based on the data and the context

Recommendation 8: Statistics assessments should make use of real data and contain authentic tasks that mirror the work of practising statisticians, supported by technology tools where possible.

Recommendation 9: Statistics assessment design should foreground conceptual work and reduce emphasis on procedures and techniques. This should begin with establishing a model of statistical thinking that identifies a progression framework of the key elements and exemplifies what each of those elements looks like at different levels.

2.5 Responsibility: ethics and data stewardship

Kathotia and Smith (2026) discuss the importance of **data stewardship** (considering the purpose, origin, and accuracy of data, and ethical use of data) in their rapid review of the statistics curriculum, building on recommendations from Jacques and Joubert (2024), and recommending that data stewardship be embedded throughout the statistics curriculum across all ages and stages. Wolff et al. (2016) include the consideration of ethics within their model of data literacy, as do Bargagliotti et al. (2020) who focus on the ethical use of data including the handling of sensitive data.

While these sources refer to data stewardship and ethics through discussion of curriculum content, responsible use of data is not currently well-represented in existing mathematics and statistics curricula in the UK. If, as evidence suggests, this aspect of data literacy should be embedded in curricula then it is also important to include in assessment despite any lack of assessment specific research in this area. Methods of assessing data stewardship will require further research, careful consideration and ongoing evaluation.

Recommendation 10: Data stewardship and ethical use of data should be embedded within statistics assessments. As a relatively innovative aspect of the mathematics and statistics curriculum, further research is necessary to design effective assessment approaches that will require ongoing evaluation to ensure construct validity.

In this section we have considered in more detail the five features of evidence-informed statistics education, and recommendations that might follow. In the final two sections we consider practice from across the four nations of the UK and from other jurisdictions in light of these features.

Section 3: Review of current approaches to assessment in the four nations of the UK

3.1 Background

The four nations of the UK have devolved educational systems, with similar but distinct approaches to learning and teaching, assessment and curriculum. England follows the National Curriculum, with National Curriculum assessments at age 10–11, GCSEs at age 15–16, and A-Levels or BTECs and the new T-Levels at age 17–18. Wales has the Curriculum for Wales, introduced in 2022, GCSEs at age 15–16, and A-Levels or BTECs at age 17–18. Scotland has the most distinct system, without GCSEs or A-Levels; instead they have National 4 and 5 assessments at age 15–16, and Highers/Advanced Highers at age 16–18; Scotland follows the Curriculum for Excellence, and has National Standardised Assessments (and their Gaelic language parallel MCNG: Measaidhean Coitcheann Nàiseanta airson Foghlam tron Ghàidhlig) at ages 4–5, 7–8, 10–11, and 13–14, although these are labelled as specifically not high-stakes, but adaptive, online, centrally-marked tests that teachers can administer at any time in the school year. Northern Ireland follows the Northern Ireland Curriculum, with GCSEs at age 15–16, and A-Levels or BTECs at age 17–18, but also has the Transfer Test, a grammar school entrance exam, at age 10–11, which is not compulsory but common.

3.2 Review and reform

Northern Ireland commissioned an Assessment Review in 2025–2026, England in 2024–2025 (Curriculum and Assessment together), Scotland 2021–2023, and Wales not officially since 2011, although they have an ongoing Qualified for the Future assessment reform centred on Made-for-Wales GCSEs that is being rolled out 2025–2027. Table 5 shows the current position in the reform cycle of each nation.

Table 5: Curriculum and assessment reform in the four nations of the UK

Nation	Last curriculum update	New curriculum plans	Last assessment review
Scotland	2010	Ongoing reform 2026-2028	Conducted 2021-2023 by Louise Hayward; published June 2023 (Hayward, 2023)
Northern Ireland	2007	For teaching 2027-2028	Conducted 2024-2025 by Lucy Crehan, published June 2025 (Crehan, 2025); although this is a curriculum review with some assessment content
Wales	2022	not known	Conducted 2011-2012 by Huw Evans, published November 2012 (Evans, 2012)
England	2014	For teaching 2027-2028	Conducted 2024-5 by Becky Francis, published November 2025 (Francis, 2025); joint curriculum and assessment review

3.3 Review of UK statistics assessment against the five features

In order to understand how statistics is assessed in the four nations of the United Kingdom, we undertook a rapid review of some of the main national qualifications and summative assessments taken by school-aged students in mathematics or statistics. Given the range of different qualifications across ages and stages, it is complex to summarise assessment practices. The following sections therefore represent a snapshot of the assessments reviewed, providing a high-level overview of how each nation typically assesses statistics. The commentary is informed by our professional judgement of the statistical content of items in the papers analysed; we have identified some common issues as well as some examples of good practice, which are illustrative rather than comprehensive or systematic. For further details of the approach taken, please see the Appendix.

3.3.1 Literacy

While specifications and guidance do refer to ideas such as **statistical thinking** and **data literacy**, in the main they are not well reflected in most of the paper-based exams. The majority of paper-based exams across the four nations use highly structured items that direct students to use particular statistical techniques or approaches. Short- and medium-item formats with highly structured mark schemes tend to limit the amount of independent reasoning and decision making available to students as key decisions are made for them by the assessment designers. It is noteworthy that even specialist Statistics qualifications take this approach in the majority of items.

In Scotland, the teacher-assessed four-unit National 4 Applications of Mathematics includes projects or activities designed to provide evidence of reasoning skills in a statistical context in one unit. Higher Applications of Mathematics includes an externally marked project (32%). These may provide more opportunities for students to be assessed on statistical thinking and/or data literacy.

A common form of paper-based exam item that attempts to assess aspects of statistical thinking asks students to critique a statement made by another person, but marks tend to be awarded for correctly identifying an appropriate or inappropriately applied technique rather than, for example, reasoning about variation. Northern Ireland's GCSE Statistics Paper 2 was notable for containing a greater proportion of literacy-related elements. However, these too tended to be short and highly directed.

3.3.2 Technology

In most qualifications across the four nations, access to technology is limited to scientific or graphing calculators only. The exception to this is Scotland: in the teacher-assessed components of Scottish qualifications, students are able to select activities or projects that can incorporate software. The project component (30 marks) of the Higher Applications of Mathematics (HAOM) mandates use of statistical software (e.g. R studio). In the HAOM paper-based exam (65 marks) students must have access to a computer and are provided with spreadsheet files with items related to modelling in both mathematics and statistics.

3.3.3 Enquiry

In the papers reviewed across all nations, while some limited references within items relate to earlier elements of the PPDAC cycle, the majority of marks are awarded for mathematical aspects of data analysis. A smaller number of marks are attributed to students' conclusions, but these are usually highly directed. The structure of marking rubrics across all qualifications was highly consistent, with all reviewed paper-based exams graded using a point-based analytic mark scheme that assigns individual marks against stages in a structured solution. This approach signals a high value on procedural work, and in the papers we reviewed results in proportionally more marks being assigned to questions that require multiple stages of mathematical working. In contrast we did not find any items on the exam papers reviewed with high marks (e.g. 5 or more) allocated for extended writing and statistical argument.

Including an extended project and internal task-based assessment in several of Scotland's qualifications suggest that in principle students have opportunities to engage in enquiry; but for the internally assessed component, this may depend on the specific tasks and activities used by teachers.

3.3.4 Authenticity

Across all four nations, the majority of paper-based exams awarded the majority of marks for **mathematical working** rather than **statistical thinking**. This was particularly apparent in items where multi-stage calculations were necessary. In post-16 qualifications this tended to be most apparent in questions related to statistical tests of significance, which were often allocated 7-9 marks, most of which were awarded for successfully executing steps in a calculation-heavy process. At earlier stages of education, longer questions tended to be associated with calculations needed to construct and read from representations such as pie charts and histograms. The majority of questions across all four nations' paper-based exams tended to be short and highly structured with limited need for contextual sense-making. Where sense-making was required, few marks were available. Marks awarded for sense-making tended to be the final 1 or 2 marks in a multi-part item and were highly predictable in format.

Across the papers reviewed for all four nations there was evidence of authentic data and contexts being used; however, this does not always extend to solving authentic statistical problems due to the highly scaffolded and directed nature of most items and the associated point-based analytic marking rubrics. Papers reviewed from

Scotland, Northern Ireland, and Wales demonstrated a greater use of cultural contexts, mostly through reference to those nations' specific national identity and cultural norms. Qualifications that included pre-release materials or additional supporting materials (e.g. A-level Mathematics and various Core Mathematics qualifications in England and GCSE Statistics in Northern Ireland) had greater opportunities for authenticity but these too were limited by approaches to assessment items and marking rubric formats that assigned the majority of marks to **mathematical working**. Alternative forms of assessment in some of Scotland's qualifications allow greater opportunities for authenticity, particularly in extended project-based components.

3.3.5 Responsibility

There was limited evidence across the papers reviewed of assessment items related to ideas of ethics and data stewardship.

3.4 Summary of assessment

In Table 6 we summarise our analysis of the current approach to assessment in the four nations of the UK with respect to the five features. While there were examples of good practice in statistics assessment across the papers and specifications that we reviewed, we did not see sufficient evidence of the five features across the full range of qualifications in any individual nation, or within individual qualifications, to recommend any nation's approach. Based on our review of the papers we argue that a key structural issue relates to the prominence of short-structured items and point-based analytic mark schemes, which appear to limit opportunities to assess statistical literacy and privilege mathematical working rather than statistical thinking when allocating marks. Access to statistical software on computers remains lacking across three of the four nations, with only Scotland requiring it in some components of assessment.

At present, including alternative forms of assessment alongside or instead of paper-based exams in some of Scotland's qualifications does create more opportunities for assessment to meet the five features. However, there are some important caveats:

1. Not all of Scotland's qualifications include alternative forms of assessment
2. The paper-based assessments still appear to overvalue mathematical working rather than statistical thinking and reasoning

3. Where teachers can choose activities for task-based internal assessment, it is beyond the scope of this review to judge how consistently these tasks might align with the five features.

Table 6: Overview of each of the four nations' current approach to statistics assessment against the five features

Recommendation	Wales	Northern Ireland	England	Scotland
Literacy				
Technology				
Enquiry				
Authenticity				
Responsibility				

No evidence or limited evidence of assessment design supporting this feature
Some evidence of assessment design that supports this feature but inconsistent, superficial, or limited to some qualifications or components
Extensive evidence of assessment design that supports this feature

Section 4: Some examples of statistics assessment in other jurisdictions

4.1 Overview

There is widespread disparity in practice in teaching and assessing statistics internationally. According to Witte et al. (2024), teaching ‘comprehensive’ data literacy (defined as including data exploration, data management, and data evaluation) is obligatory in four EU countries: Ireland, the Netherlands, Austria, Lithuania and partially addressed in Germany and Finland; other countries address aspects of data literacy such as collecting and interpreting data, but do not use the term ‘data literacy’. Countries such as Germany, Turkey, Ireland, and Cyprus are suggested to emphasise data-driven decision-making in statistics at primary level. It is claimed that the highest level of statistical literacy is taught in countries like Aotearoa New Zealand and Japan, where simulation-based methods and technology are well embedded in the curriculum (Burrill, 2023).

Jones and Dudzic reviewed statistics education practices in the UK, suggesting that the UK integrates statistics into mathematics much more than countries such as the US, which opts for more standalone, specialist courses. Other issues include the lack of real-world scenarios in UK statistics education: ‘One criticism of statistics education in the UK lies in the lack of emphasis on practical applications and real-world scenarios. In contrast to the USA and NZ, where there is a greater focus on applying statistical methods to solve authentic problems, the UK's curriculum may be perceived as more theoretical.’ (Jones & Dudzic, 2024, p. 1).

4.2 Aotearoa New Zealand

In Aotearoa New Zealand, the 2007 curriculum includes Mathematics and Statistics, rather than Mathematics alone. The curriculum was reformed in 2025 with a phased implementation between 2026 and 2030. At the time of this review relevant assessment materials were not available for the new curriculum, so the following section refers to currently available documentation based on the 2007 curriculum.

The curriculum documentation states ‘Mathematics is the exploration and use of patterns and relationships in quantities, space, and time. Statistics is the exploration and use of patterns and relationships in data.’ The statistical enquiry process is

explicitly referenced, and students move through learning to use it to using it 'with statistical justification' and then 'with statistical insight'

In Aotearoa New Zealand in Years 3–8 (age 7–13), students have compulsory twice-yearly standardised testing, using a government-approved tool, such as PATs (Progressive Achievement Tests; described as 'formative, low-stakes assessments') or SMARTs (Student Monitoring, Assessment and Reporting Tool). In Year 10 most students take the NCAA (Numeracy Common Assessment Activity) a threshold assessment that they can take multiple times until they pass.

Older students in Aotearoa New Zealand can take their NCEA (National Certificate of Educational Achievement, run by NZQA) Level 1 in Year 11, Level 2 in Year 12 and Level 3 in Year 13. These are qualifications with both externally-assessed and internally-assessed exams, the latter marked by teachers and moderated by NZQA. One aspect of the assessment specifically focuses on the objective of 'Explore data using a statistical enquiry process' and students are able to choose the direction of their own investigation, including investigative question or statement, with assessor approval. Student assessment outcomes are: Not Achieved, Achieved, Merit or Excellence. Students write about their statistical enquiry using headings such as:

- Purpose
- Prediction
- Plan
- Variation
- Analysis (centre, spread, shape, unusual features)
- Conclusion

In particular, the focus on students using the whole statistical enquiry cycle and highlighting variation as a key aspect is considered to be excellent practice.

In addition to our review of UK assessment described in the previous section, we also reviewed Aotearoa New Zealand's assessment of Mathematics and Statistics at Levels 1, 2, 3, and Scholarship Statistics.

4.3 The USA: Levels of Conceptual Understanding in Statistics Project (LOCUS)

LOCUS was a large-scale, US-based research project that produced statistics assessment items designed to measure conceptual understanding across the statistical problem-solving process. The items ask students to demonstrate statistical reasoning with minimal calculation or graph construction based items

(Jacobbe et al., 2023). The LOCUS assessments have been used with students in US grades 6–12 (age 11–18) as well as with both trainee and qualified teachers.

LOCUS items were designed to reflect the recommendations for curriculum and assessment made in the original GAISE report (Franklin et al., 2007). They contain a mixture of multiple-choice and constructed response items across three levels of sophistication described in the GAISE framework. The project makes sample items available at <https://locus.thinkdataed.org/>

As a further addition to our review of assessment, we also analysed sample items from the LOCUS project at grades 6, 7, 8, and 9–12 as examples of assessment that operates within the restrictions of paper-based exam structures while also assessing statistical thinking.

Table 7: Overview of Aotearoa New Zealand and LOCUS (US) approaches to statistics assessment against the five features

Recommendation	Aotearoa New Zealand	LOCUS (US)
Literacy		
Technology		
Enquiry		
Authenticity		
Responsibility		

No or limited evidence of assessment design supporting this feature
Some evidence of assessment design that supports this feature but inconsistent, or limited to some qualifications or components
Extensive evidence of assessment design that supports this feature

Both sets of paper-based assessment materials reviewed have an explicit focus on statistical thinking across the items. Marking rubrics support this by using levels of response. The two take different approaches: the Aotearoa New Zealand assessments use exemplar responses at different levels of competency, while the LOCUS items exemplify a model correct solution for each sub-item and award between zero and four marks for the whole item based on the combination of sub-

items judged fully correct, partially correct, or incorrect. The LOCUS materials also include multiple-choice questions that still demand students engage in statistical reasoning.

In Aotearoa New Zealand, several curriculum standards are assessed internally through tasks and activities and students have access to software during these components. For the LOCUS items a calculator may be used; however, few items require students to compute. The LOCUS assessments can be taken as either paper-based or online tests. However, the online version is restricted to multiple-choice items and does not require use of statistics software, despite being computer-based. This makes LOCUS a useful exemplar for UK assessment designers seeking to assess statistical problem-solving where resource limitations could mean access to computers in assessments is not possible at scale in the short-term. This approach does not align well with feature 2 (technology) and therefore should not be treated as an effective way to assess statistics in the longer term unless used alongside other forms of assessment that support the use of statistical software.

Both sets of materials foreground the PPDAC cycle and attempt to prompt authentic statistical enquiry and problem-solving. The LOCUS assessments represent one approach to creating a more balanced approach to assessing the stages of the statistical problem-solving cycle. LOCUS assessments contain short- and medium-structured items that capture individual elements of the statistical problem-solving cycle, with assessments designed to ensure a balance of items across four domains of statistical problem-solving: Formulating statistical questions (10–20%); Collecting Data (20–30%); Analysing Data (25–35%); and Interpreting Results (25–35%).

While not fully meeting the principles laid out in the five features of evidence-informed assessment, these two sets of materials include examples of various approaches to statistics assessment that could be modified or incorporated by Northern Ireland, Wales, England, and Scotland to improve on their current models.

Conclusion

In this report we set out some organising principles for statistics education and considered previous reports from the Joint Mathematical Council of the United Kingdom, the Royal Society, and the Royal Statistical Society, using them to establish five features of evidence-informed statistics assessment. Based on a rapid review of literature we exemplified the five features with reference to selected relevant

research and made ten recommendations for policy makers supported by the available evidence.

We reviewed a range of assessments from across the four nations along with two key international examples from Aotearoa New Zealand and the US. In this review we attempted to demonstrate how closely aligned current assessment design was to the five features, identifying limitations and barriers, as well as some examples of good practice. Currently, none of the four nations' assessments are well aligned with any of the five focuses, although there are some examples of good practice across some of the four nations that could be built on. Three limitations emerged from our review that need to be addressed in order enhance alignment between assessments and the five features:

1. Access to statistical software within assessments.
2. Paper-based exams limit opportunities for students to conduct investigative work and engage with the full PPDAC cycle.
3. Short-structured items and point-based analytic mark schemes reward **mathematical working** at the expense of **statistical thinking**.

In order to act on the recommendations made in this report, addressing these limitations will be an important first step.

References

- Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)*. NCTM.
- Bargagliotti, A., & Groth, R. (2016). When mathematics and statistics collide in assessment tasks. *Teaching Statistics*, 38(2), 50–55.
<https://doi.org/10.1111/test.12096>
- Ben-Zvi, D., & Garfield, J. (2004). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges. In D. Ben-Zvi & J. Garfield (Eds), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 3–15). Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_1
- Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of Statistics Learning Environments. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds), *International Handbook of Research in Statistics Education* (pp. 473–502). Springer International Publishing. https://doi.org/10.1007/978-3-319-66195-7_16
- Ben-Zvi, D., Makar, K., & Garfield, J. (Eds). (2018). *International handbook of research in statistics education*. Springer International Publishing.
<https://doi.org/10.1007/978-3-319-66195-7>
- Bidgood, P. (2010). Relating Assessment to the Real World. In *Assessment Methods in Statistical Education* (pp. 153–161). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9780470710470.ch12>
- Biggs, J. B., & Collis, K. (1982). *Evaluating the quality of learning: The Solo Taxonomy*. Academic Press.
- Burkhardt, H., & Schoenfeld, A. (2018). Assessment in the service of learning: Challenges and opportunities or plus ça change, plus c’est la même chose. *ZDM*, 50(4), 571–585. <https://doi.org/10.1007/s11858-018-0937-1>
- Burrill, G. (2023). An International Look at the Status of Statistics Education. In G. Burrill, L. de Oliveria Souza, & E. Reston (Eds), *Research on Reasoning with Data and Statistical Thinking: International Perspectives* (pp. 11–16). Springer International Publishing. https://doi.org/10.1007/978-3-031-29459-4_2

- Cao, L. (2017). Data Science: A Comprehensive Overview. *ACM Computing Surveys (CSUR)*, 50(3), 43:1–43:42. <https://doi.org/10.1145/3076253>
- Chan, S. W., & Ismail, Z. (2014). A Technology-Based Statistical Reasoning Assessment Tool in Descriptive Statistics for Secondary School Students. *The Turkish Online Journal of Educational Technology*, 13(1).
- Chan, S. W., Ismail, Z., & Sumintono, B. (2015). Assessing statistical reasoning in descriptive statistics: A qualitative meta-analysis. *Jurnal Teknologi*, 72(2). https://knova.um.edu.my/research_publications_2011_2015/5510
- Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910677>
- Cobb, G. W. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (pp. 3–33). Mathematical Association of America.
- Crehan, L. (2025). *A Strategic Review of the Northern Ireland Curriculum*. https://www.education-ni.gov.uk/sites/default/files/2025-06/NI%20Curriculum%20Review_0.pdf
- Cui, Y., Chen, F., Lutsyk, A., Leighton, J. P., & Cutumisu, M. (2023). Data literacy assessments: A systematic literature review. *Assessment in Education: Principles, Policy & Practice*, 30(1), 76–96. <https://doi.org/10.1080/0969594X.2023.2182737>
- Davies, N., & Marriott, J. (2010). Assessment and Feedback in Statistics. In *Assessment Methods in Statistical Education* (pp. 1–19). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470710470.ch1>
- 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>
- delMas, R. C. (2002). Statistical Literacy, Reasoning, and Learning: A Commentary. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910679>

- Dichev, C., & Dicheva, D. (2017). Towards Data Science Literacy. *Procedia Computer Science, International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland, 108*, 2151–2160.
<https://doi.org/10.1016/j.procs.2017.05.240>
- Evans, H. (2012). *Review of Qualifications for 14 to 19-year-olds in Wales Final report and recommendations*.
https://dera.ioe.ac.uk/id/eprint/16082/7/121127reviewofqualificationsen_Redacted.pdf
- Francis, B. (2025). *Curriculum and Assessment Review final report: Building a world-class curriculum for all*.
https://assets.publishing.service.gov.uk/media/690b96bbc22e4ed8b051854d/Curriculum_and_Assessment_Review_final_report_-_Building_a_world-class_curriculum_for_all.pdf
- Franklin, C. A., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. L. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. American Statistical Association. <http://www.professores.im-uff.mat.br/hjbortol/disciplinas/2012.2/esp00001/arquivos/gaise-1.pdf>
- Gal, I., & Garfield, J. B. (1997). *The Assessment Challenge in Statistics Education*. IOS Press.
- Garfield, J., & Franklin, C. (2011). Assessment of Learning, for Learning, and as Learning in Statistics Education. In C. Batanero, G. Burrill, & C. Reading (Eds), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (Vol. 14, pp. 133–145). Springer Netherlands.
https://doi.org/10.1007/978-94-007-1131-0_16
- Gibeau, R.-M., & Cousineau, D. (2025). Uncovering key statistical concepts from an investigation of four standardized tests. *Teaching Statistics*, 47(3), 219–234.
<https://doi.org/10.1111/test.12410>
- Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1), 22–25. <https://doi.org/10.52041/serj.v16i1.209>

- Hayward, L. (2023). *It's Our Future – Independent Review of Qualifications and Assessment: Report*. <https://www.gov.scot/publications/future-report-independent-review-qualifications-assessment/>
- Jacobbe, T., delMas, B., Hartlaub, B., Haberstroh, J., Case, C., Foti, S., & Whitaker, D. (2023). Establishing the Validity and Reliability of the LOCUS Assessments. *Numeracy*, 16(1). <https://doi.org/10.5038/1936-4660.16.1.1420>
- Jacques, L., & Joubert, M. (2024). *A vision for meaningful data education across the four nations of the UK*. JMC. <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 statistics curriculum in the UK*. <https://rss.org.uk/RSS/media/File-library/Policy/2024/Recommendations-for-statistics-curriculum-in-the-UK-full-paper.pdf?ext=.pdf>
- Kathotia, V., & Smith, C. (2026). *Data education and mathematics 5–16: A rapid review for the Joint Mathematical Council of the UK and the Royal Society*. Open University. <https://www.jmc.org.uk/wordpress-cms/wp-content/uploads/2026/02/Rapid-Review-of-Data-Education-within-Mathematics-5-16-FINAL.pdf>
- Macey, D., & Rycroft-Smith, L. (2022). The keys to the kingdom: Why we can't ignore assessment if we care about real improvement in the teaching and learning of statistics. *Bridging the Gap: Empowering and Educating Today's Learners in Statistics: Proceedings of the 11th International Conference on Teaching Statistics*. ICOTS11. https://fpce.uc.pt/iase-web/icots/11/proceedings/pdfs/ICOTS11_144_MACEY.pdf
- Macgillivray, H. (2010). Variety in Assessment for Learning Statistics. In *Assessment Methods in Statistical Education* (pp. 21–34). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470710470.ch2>
- Porkess, R. (2013). A world full of data. *Statistics Opportunities across A-Level Subjects*. London: Royal Statistical Society/The Institute and Faculty of Actuaries. <https://www.actuaries.org.uk/documents/world-full-data-statistics-opportunities-across-level-subjects>

- Ridgway, J., & Nicholson, J. (2019). Problematising high-stakes assessment in statistics. In J. M. Contreras, M. M. Gea, M. M. Lopez-Martin, & E. Molina-Portillo (Eds), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*.
https://digibug.ugr.es/bitstream/handle/10481/55044/ridgway_nicholson.pdf?sequence=1&isAllowed=y
- Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. *International Electronic Journal of Mathematics Education*, 14(3), 535–546.
<https://doi.org/10.29333/iejme/5755>
- Saidi, S. S., & Siew, N. M. (2022). Assessing Secondary School Students' Statistical Reasoning, Attitude Towards Statistics, and Statistics Anxiety. *Statistics Education Research Journal*, 21(1), 6–6. <https://doi.org/10.52041/serj.v21i1.67>
- The Royal Society. (2024). *A new approach to mathematical and data education*. The Royal Society.
- Witte, V., Schwering, A., & Frischmeier, D. (2024). Strengthening Data Literacy in K-12 Education: A Scoping Review. *Education Sciences*, 15(1), 25.
<https://doi.org/10.3390/educsci15010025>
- Wolff, A., Gooch, D., Caverio 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). <https://doi.org/10.15353/joci.v12i3.3275>

Appendix: Approach to review of assessments

For each included assessment we looked at the specification document and the most recent full set of assessment materials (assessment, marking rubrics, and any additional materials). Most sets of materials are taken from Summer 2025 as for many qualifications the marking criteria are withheld for the most recent assessment cycle. For GCSE in Wales we used the published sample assessment materials for a new Mathematics and Numeracy double award due for first assessment in 2027.

Table A1: Framework for analysis

Document type	Data extracted
Specification (or qualification web pages where no specification available)	• Format of assessment components • Location of statistics content within the assessment • Whether statistics is optional or mandatory within the qualification • Access to technology • Existence of pre-release or supplementary materials
Exam paper	• Form and format of items • Confirmation of location of statistics within assessment • Confirmation of access to technology
Additional materials	• Content of pre-release or supplementary materials
Marking guidance document	• Structure of marking rubric • Confirmation of item form and format

For each qualification we read the assessment schedule in the specification or guidance page and recorded the details. We then read through each available document in the assessment materials in order of additional materials, question paper, and marking guidance. We used an inductive approach to generate simple

descriptive codes for each category of data and recorded them in a spreadsheet. The recorded codes do not capture any year-on-year variety in assessment materials and so represent a snapshot of the qualification in the year the materials were published. Codes generated for item format and marking guidance represent our professional judgement based on interpreting the main features of each item.

Below is an illustration of exemplar questions we asked when making judgements about each of the five focuses:

1. Literacy

- Did the format of the assessment create space for extended statistical thinking and/or data literacy?
- Did the assessment allow students to make meaningful decisions in relation to statistical information?
- Did the approach to marking award credit for statistical thinking?

2. Technology

- Did students have access to statistical software/technology tools in the assessment?
- How were students required to use technology tools in the assessment?
- Did the technology used in the assessment allow students to make meaningful statistical decisions?

3. Enquiry

- Did the assessment provide opportunities for students to engage with the full PPDAC cycle?
- How balanced was the treatment of each element of the PPDAC cycle across the assessment?
- Did the approach to marking award reasonable credit for engaging with all elements of the PPDAC cycle?

4. Authenticity

- Did the assessment incorporate real or realistic contexts and data?
- Were the majority of marks awarded for contextual sense-making rather than mathematical working?
- Did the assessment require students to solve genuine statistical problems?

- Did the assessment incorporate contexts related to local and/or global cultural identities?

5. Responsibility

- Was there evidence in the assessment of students being asked to consider data provenance?
- Were there questions that required students to consider ethical use of data?

The following descriptions were written to capture the intended meaning of specific codes.

Short-structured item format – short, structured items, often scaffolded into multiple parts. usually closed questions targeted at specific curriculum content statements. Usually 1–4 marks per part

Medium-structured item format – slightly longer, structured items, usually closed questions, usually targeted at specific curriculum content statements but requiring extended working. Usually 5–9 marks. Often sub-parts are preceded by short-structured items.

Extended response format – longer, less structured items. Usually more open questions. Usually more holistic encompassing multiple curriculum content statements and/or broader competencies.

Point-based analytic mark scheme – highly structured rubric where marks are allocated to specific actions or values within in a model solution.

Levels-of-response rubric – evaluative rubric where a level is allocated to the complete solution based on exemplification of an ideal response at each.

Hybrid – a rubric that contains both points based elements alongside levels of response, for example marking each sub-item through levels of response and then assigning overall marks based on combining the assigned levels in each sub-item, or a points-based analytic rubric where a small number of marks within an item are assigned based on levels-of-response.

Table A2 Papers included in the assessment review

England	KS2 National Curriculum Test papers • https://www.gov.uk/government/publications/key-stage-2-tests-2026-mathematics-test-materials KS4 June 2025 Mathematics GCSE papers • AQA GCSE (9-1) Mathematics (8300) • Pearson Edexcel GCSE (9-1) Mathematics (1MA1) • OCR GCSE Mathematics (9-1) (J560) June 2025 Statistics GCSE papers • AQA GCSE (9-1) Statistics (8382) • Pearson Edexcel GCSE (9-1) Statistics (1ST0) KS5 June 2025 A-level mathematics papers • AQA A-level Mathematics (7357) • OCR Mathematics A (H240) • OCR Mathematics B (MEI) (H640) • Pearson Edexcel A-level Mathematics (9MA0) June 2025 A-level further mathematics • OCR Further Mathematics A (H245) • OCR Further Mathematics B (MEI) (H645) • Pearson Edexcel A-Level Further Maths (9FM0) A-level statistics • June 2024 Pearson Edexcel A-Level Statistics (9ST0) Core Mathematics June 2025 • AQA Level 3 Certificate in Mathematical Studies (1350) • OCR Core Maths A (MEI) Level 3 Certificate (H868) • OCR Core Maths A (MEI) Level 3 Certificate (H869) • Pearson Edexcel Mathematics in Context (Level 3 Core Maths) (7MC0)
Wales	KS4 Sample assessment materials (first assessment 2027) – final first assessment in current specs has passed • WJEC GCSE Mathematics and Numeracy (double award) KS5 • WJEC AS/A-level Mathematics Summer 2025 • WJEC AS/A-level Further Mathematics Summer 2025
Northern Ireland	KS4 • CCEA GCSE Mathematics (2017) Summer 2025 • CCEA GCSE Statistics (2017) Summer 2025

	KS5 • CCEA GCE AS/A-level Mathematics Summer 2025 • CCEA GCE AS/A-level Further Mathematics Summer 2025
Scotland	National 4 Mathematics • SQA Guidance pages https://www.sqa.org.uk/sqa/47417.html National 5 Mathematics • SQA National 5 Mathematics May 2026 National 3 Applications of Mathematics • SQA Guidance pages https://www.sqa.org.uk/sqa/81287.html National 4 Applications of Mathematics • SQA Guidance pages https://www.sqa.org.uk/sqa/81279.html National 5 Applications of Mathematics • SQA National 5 Applications of Mathematics May 26 Higher Applications of Mathematics • SQA Higher Applications of Mathematics May 26 Advanced Higher Statistics • SQA Advanced Higher Statistics May 26 Statistics Award (SCQF level 6) • SQA Guidance pages https://www.sqa.org.uk/sqa/73224.html
Additional papers	Aotearoa New Zealand Mathematics and statistics https://www2.nzqa.govt.nz/ncea/subjects/past-exams-and-exemplars/mathematics-and-statistics/#e12274 • Level 1, 2, 3, and Scholarship Statistics 2025 past papers. LOCUS project https://locus.thinkdataed.org • Grades 6,7,8,9 exemplar items