Department for Education

Principles for evidence-informed policymaking in fast-changing technology contexts

September 2026

Authors: Department for Education Science Advisory Council – AI and Digital Education working group

Contents

Executive summary

Technology in education, particularly AI, is moving faster than traditional evaluation methods can accommodate. The pace of change has challenged previous approaches to evaluation in education, such as randomised controlled trials, which have previously provided evidence for policymaking and advice to educators and parents.

We propose here eight principles as a framework to support evidence-informed policymaking in fast-changing technology contexts. This approach aims to synthesise all available evidence to fully understand a topic but will also continually update our understanding as new evidence arrives. Policymaking within this framework triangulates imperfect evidence, updates judgements over time, matches methods to purpose, monitors harm and equity, preserves flexibility, evaluates systems and products, and makes evidence accessible to users.

This framework was developed by the Department for Education (DfE) Science Advisory Council's AI and Digital Education Working Group to support policymaking where technology evolves more quickly than traditional research cycles can provide definitive evidence, e.g. artificial intelligence (AI) in education.

The eight principles

The logic of evidence

1. Combine and triangulate evidential sources, treating each as partial: No single source of evidence is sufficient in rapidly changing contexts; policymakers should draw on and triangulate multiple evidence sources, recognising that each contributes only part of the overall picture.
2. Match evidence to purpose: Evidence requirements should be tailored to the question being addressed, recognising that product improvement, practitioner guidance and harm detection each require different methods, timescales and standards of proof.
3. Treat evidence collection as ongoing and formative: Evidence gathering should be a continuous process that updates understanding over time, enabling decisions to adapt as new information emerges and technologies evolve.

Methods of evidence generation

4. Keep options open: Policy and procurement decisions should preserve future flexibility and avoid creating irreversible commitments before sufficient evidence about long-term impacts is available.

5. Evaluate conditions alongside products: Assessment should consider not only whether a particular product works, but also the wider conditions, contexts and implementation factors that influence its effectiveness.
6. Match the method to the intervention's maturity: Evaluation approaches should reflect the stage of development of a technology, using exploratory and formative methods for emerging tools and more formal evaluation as products become stable and better understood.

Infrastructure for evidence in practice

7. Build infrastructure that supports evidence accumulation: Governments, providers and institutions should create systems that enable continuous evidence collection, including access to usage data and structured mechanisms for sharing practitioner experience.
8. Make evidence accessible to its intended audiences: Evidence should be translated into clear, practical and audience-appropriate formats so that policymakers, educators, school leaders, parents and other users can apply it effectively.

About this framework

This document was produced by the AI and Digital Education working group of the Science Advisory Council (SAC) as independent advice to the Department for Education (DfE). The SAC and its working groups provide independent scientific advice to the DfE on matters relevant to policy development and delivery.

This framework was developed to support DfE in evaluating the existing evidence base on the use of AI in education, but the principles set out below apply more widely to any policy context where the technology under consideration changes faster than the conventional research cycle and where waiting for definitive evidence would mean not providing guidance to schools and families for an extended period. The authors emphasise that conventional approaches remain appropriate where evidence cycles match decision cycles, or where the precautionary principle can be applied with confidence.

This framework is intended to serve two audiences:

- Inside government, it offers a shared reference for evidence-informed decisions about AI in education and a structure for explaining those decisions to Ministers and the public.
- Outside government, it sets out the evidential standards about AI in education that the SAC believe should be met, and the kinds of questions that developers and providers should be prepared to answer.

This framework is not intended to replace existing DfE programmes and guidance in this space. It sits alongside and is consistent with existing DfE activities as well as DfE's broader strategy.

A glossary of terms in bold is provided.

Contributors

Professor Dame Athene Donald DBE, FRS– Professor Emerita of Experimental Physics, University of Cambridge and Chair of the DfE Science Advisory Council

Professor Rose Luckin CBE – Professor Emerita of Learner Centred Design, University College London, and Member of the DfE Science Advisory Council

Professor Mark Mon-Williams FRCPCH – Chair of Cognitive Psychology, University of Leeds, and Deputy Chair of the DfE Science Advisory Council

Professor Amy Orben – Programme Leader of the Digital Mental Health Research Programme, MRC Cognition and Brain Sciences Unit, University of Cambridge, and Member of the DfE Science Advisory Council

Professor Michael Thomas – DfE Chief Scientific Adviser and Professor of Cognitive Neuroscience at Birkbeck University of London

Summary: How we can make good decisions about new technology

This document is about how governments and schools can make good decisions about fast-changing technology, especially artificial intelligence (AI). AI changes so quickly that by the time researchers have finished a study and published the results, the technology may already be very different. Because of this, waiting for one perfect piece of evidence is often not the best way to make decisions.

Instead, the report argues that people should collect information from many different sources and keep updating what they know as new evidence becomes available. Instead of looking for one final answer, they should build the best understanding they can using all the information they have, and tools should be developed to help them do that.

The authors created eight principles to help people make decisions about new technology. These principles can be used for AI in education and for other technologies that change very quickly.

The first principle is to use different types of evidence because no single study can tell us everything. The second principle is that different questions need different kinds of evidence. For example, finding out whether a product works, helping teachers and parents, and identifying safety risks may all need different approaches.

The third principle is to keep collecting evidence over time. As technology changes, our understanding should change too. This is similar to how teachers use regular assessments to see how pupils are progressing.

The fourth principle is to keep options open. Schools and governments should avoid making choices that are difficult to change later, especially when they still do not know enough about a technology's long-term effects.

The fifth principle is to look not only at the technology itself but also at how it is used. A tool might work well in one school but not in another, depending on things such as teacher training, the curriculum, and how closely adults supervise its use.

The sixth principle is to choose evaluation methods that fit the stage of development of the technology. New and changing tools may need feedback from teachers, experts, and users before larger studies are possible.

The seventh principle is to build systems that help collect evidence over time. For example, schools should be able to see how often AI tools are used and whether they are helping pupils. Teachers should also have ways to share their experiences with others.

The eighth principle is that evidence should be easy to understand. Research should be turned into clear and practical advice so that teachers, school leaders, parents, and policymakers can use it.

The report also says that fairness is very important. People should always check whether technology works differently for different groups of children. If it does, technology could accidentally make inequalities worse.

Overall, the report argues that when technology changes quickly, good decisions come from carefully combining different kinds of evidence, continually learning from new information, watching for possible harms, and making sure everyone can understand the evidence being used.

Introduction

The problem

Traditional research timelines are incompatible with the pace of change in AI. A '2023' or '2024' publication typically draws on data collected in 2021 or 2022. Thus, usage, capabilities, and products have often changed materially by the time research findings are published.

This problem is not new, and it is not unique to AI. Hamelink[1] set out five options available to policymakers confronting fast-moving technologies and argued that these fail when technology changes faster than the evidence cycle, leaving only summative evaluation as an option. Collingridge[2] showed that even summative evaluation is often inadequate. A technology's unwanted consequences are typically too entrenched to reverse without serious disruption, expense, and political difficulty by the time they are apparent. The Collingridge dilemma is that we cannot predict harm early enough to act efficiently, but we cannot act efficiently once the harm has become apparent. Stegenga[3] describes science conducted faster than routine practice would normally allow as 'fast science' and proposes principles for assessing when accelerated approaches are justified, which connect directly to the framing offered here.

Balancing precaution and evidence

A common academic response to the Collingridge dilemma is to suggest the application of the **precautionary principle** by default, restricting a technology until it can be shown not to cause harm.

The precautionary principle is an essential tool for policymakers, and this framework supports this principle whilst providing suggestions on how and when the principle is best applied. The question is when it should be the default and when it should be applied selectively. Where there is credible risk of serious harm to children, where consequences may be irreversible, or where the evidence required to dispel uncertainty cannot be gathered in time, precautionary restriction can be the right response. The DfE's recent guidance on under-five screen use is one example of the precautionary principle being applied in a considered way. Where those conditions do not hold, and particularly in lower-risk settings, applying precaution as a default can hold back benefits that would otherwise reach pupils and teachers, and that delay is itself a cost. The framework is concerned with this second category of case.

Applying the precautionary principle well requires careful judgement about where and when it is the right tool, and policymakers reasonably look to high-quality evidence to support that judgement. In medicine, the **Randomised Controlled Trial** (RCT) has proved to be a powerful methodology for assessing the efficacy of pharmacological and technological interventions. The success of the RCT in clinical settings has led some to assume that an RCT is always appropriate and provides the best evidence in all situations. The history of science shows that this is demonstrably false.[4] The RCT is a useful methodology and has been applied with great success in education, for example in underpinning the Teaching Toolkit produced by the Education Endowment Foundation.[5] However, it is one of several methodologies in the scientist's toolbox. While RCTs have the advantage of control, rigour, and a direct window on causality, the time required to design and execute them can mean that, in fast-changing technology contexts, the product or version evaluated has changed materially by the time results are available. This can limit extrapolation between versions or products, but it does not make the evidence redundant: where tools are relatively stable, where the underlying learning mechanism is demonstrated, or where updates are minor, RCTs can continue to provide valuable evidence alongside broader approaches to evidence generation and assurance.

Our considered opinion is that policymaking within the 'AI in education' context benefits from a wider repertoire than the choice between precautionary restriction and RCT-standard proof alone. A **Bayesian[a]** framing, in which the best current picture is built from multiple sources, sits alongside those tools rather than replacing them. This picture can then be used to determine where the precautionary principle should be applied and where a more accumulative, formative approach is warranted. In some areas, particularly those touching on potential serious harm to children, precaution is appropriate. In others, including tools for teachers that support with administrative tasks, a more accumulative approach is warranted. RCTs operating in slower time could still enable insights into the broader effects of classes of technology rather than specific products.

The empirical picture

The picture painted by empirical studies of 'AI in education' reinforces the case for a different approach. A recent Stanford meta-analysis of approximately 800 papers on 'AI in education' found only 20 studies that met the threshold for causal evidence.[6] An analysis of current UK-funded 'AI in education' research found that not one of the reviewed projects used a randomised controlled trial as its primary evaluation approach.[7] The Education Endowment Foundation, which has traditionally been associated with RCT-type evidence, has begun to recognise the need to embrace alternative methods when appropriate.[8,9]

[a] Bayes was a mathematician who pioneered an approach to probability theory where the odds are not fixed, but are continually updated as new data come in (e.g., the odds of a horse winning a race are updated when the jockey is known, the firmness of the ground becomes clear, etc.)

The organising logic: Bayesian accumulation

The principles below are organised around the concept of Bayesian Evidence Accumulation (BEA) as their common logic. The term is used here to describe the working group's organising approach, drawing on the broader Bayesian tradition of probabilistic reasoning under uncertainty rather than on a single canonical framework.

In plain terms, BEA builds up the best current picture from many imperfect sources of information, updates that picture as new information arrives, and acts on the picture as it stands rather than waiting for one definitive answer that may not arrive in time. A picture built from multiple inputs (even if those inputs provide less definitive answers), properly triangulated and updated as conditions change, is more useful in a fast-moving technology context than a single study whose findings are already dated by the time it reaches policy.

The BEA approach is consistent with the way in which expert judgement has been used in other fast-moving contexts. Expert councils have previously made consequential recommendations under significant uncertainty and without waiting for ideal data. This approach is not without error, and speed alone does not validate a method. Bayesian accumulation does not mean lower standards; it means combining inputs honestly, updating openly, and making assumptions explicit.

A note on equity

Equity is an essential cross-cutting principle that must be considered when determining the impact of technology. Any evidence framework that does not explicitly ask whether a tool, approach, or policy works differently for pupils with those demographic characteristics known to increase the risk of disadvantage (for example, special educational needs, sex, socio-economic position, ethnicity) could have the unintended consequence of increasing inequality. Every principle below should be read with this constraint in mind.

Eight principles

Eight principles are presented under three thematic groups: the underlying logic of evidence (Principles 1 to 3), the methods of evidence generation (Principles 4 to 6), and the infrastructure that makes the framework work in practice (Principles 7 and 8).

The logic of evidence

Principle 1: Combine and triangulate evidential sources, treating each as partial

In fast-changing technology contexts, the goal of evidence generation is to maintain the best current picture and update understanding as new signals arrive, rather than to deliver a final verdict.

No single source or methodology is sufficient. Reliable signals come from combining several input types, which will be weighted depending on the policy question being asked.

Quality thresholds apply to each input, which need to be comprehensively considered. Inputs of different strengths can be combined, with weaker inputs informing direction and stronger inputs anchoring conclusions. The triangulation should be designed to capture differential effects across pupil groups wherever possible (see the equity note above), even where the immediate policy decision is based on average effects.

Principle 2: Match evidence to purpose

Different evidential questions require different methodological designs. The guiding principle is to design evidence with its intended use in mind, rather than assuming that all evidence should take the same form or be judged against a single standard.

Three distinct purposes can be identified:

- improving products, a technical and commercial development question
- advising practitioners and parents, a policy and practice question,
- identifying and mitigating harm, a safety and safeguarding question.

Each purpose requires different evidential standards, timelines, and communication approaches.

Harm detection has a special status when products are used by children. A product that produces positive learning effects at a population level but causes serious harm to a subset of users cannot be defended on net-benefit grounds. Harm detection therefore needs methods that pick up serious effects on small groups of users, not only average outcomes, and should be built into evaluation from the start rather than added once products are in use. **Cognitive offloading** and erosion of independent judgement, where short-term performance gains during use can mask longer-term harm to learning when a tool is removed, are particularly clear examples of the kind of harm that needs consideration and evaluation from the outset.

Principle 3: Treat evidence collection as ongoing and formative

Evidence about 'AI in education' should be gathered continuously and adaptively, rather than as a sequence of point-in-time studies. Just as formative assessment tracks a pupil's progress over time to guide next steps, evidence about AI in education should be gathered as an ongoing process that informs decisions as they arise.

An **inferential risk** lens is useful. The cost of acting on a false positive (adopting a tool that does not work) is not the same as the cost of acting on a false negative (failing to adopt a tool that would help or failing to identify a harmful one). Policymakers should be explicit about which error is more costly, and in which context, because this shapes the burden of proof required before action.[10]

Methods of evidence generation

Principle 4: Keep options open

Policy choices should preserve flexibility rather than entrench commitments at early stages of technology adoption. The cost of changing course is low while a technology is unsettled and rises sharply once it is embedded.[2] Procurement structures that create monopolies, proprietary lock-in, or dependence on a single provider before evidence is available should be treated with particular caution.

Principle 5: Evaluate conditions alongside products

Specific products will change materially within months, but the conditions under which AI is used in education change more slowly. Evaluation should look at both aspects.

Product-level evaluation, including the work of the EdTech Impact Testbeds, has clear value. It identifies which tools work and helps surface the features that distinguish products that work from those that do not. The BEA framework supports continued investment in this work.

Evaluation should also look at the conditions that shape whether products work in practice: the quality of teacher training, the deployment context, the curriculum fit, data governance arrangements, and the degree of human oversight. Asking not just whether something works, but for whom, in what circumstances and through what underlying process (also called a **Context-Mechanism-Outcome lens**) is useful. The mechanism that delivers a learning gain is rarely the branded product itself; it is some broader feature of the interaction between the product, the user, and the setting. Identifying such mechanisms and communicating them to stakeholders, so they understand the important properties of technology beyond specific products, should be a high priority for evaluations.

Principle 6: Match the method to the intervention's maturity

The appropriate evidential methodology depends on the stability and understanding of the AI application. For novel or rapidly changing tools, the sequence should be as follows:

- define risk boundaries and use context
- gather practitioner experience and expert input
- triangulate practitioner experience and expert input with lived experiences (which constitute a different but valid form of evidence)
- design more formal evaluation once the intervention has stabilised

The demand for RCT-standard evidence before practitioner guidance is feasible is not a neutral methodological choice. In a fast-moving technology context, it functions as a decision to withhold guidance indefinitely.

Infrastructure for evidence in practice

Principle 7: Build infrastructure that supports evidence accumulation

Bayesian, formative evidence accumulation requires infrastructure that does not yet fully exist. Two elements are particularly important.

The first element is transparency of usage data. Longer-term evaluation of AI tools, including evaluation across product iterations, depends on usage data being accessible to schools and to independent evaluators. If schools can see how much pupils and teachers use a product, they can assess engagement, equity of use, and link patterns to outcomes such as attainment over time. One way to do this is to require that product developers provide a consistent set of usage data indicators in an accessible format, with an agreed minimum dataset, as a condition of public sector procurement. The onus would thus be on the provider to demonstrate, through ongoing provision of this data, that the product is safe and effective in use (which can then be independently assessed). The exact specification would be developed alongside relevant DfE teams and with developers themselves. This infrastructure also supports a clear distinction between developers' internal product evaluation, aimed at improving the product, and independent external evaluation, aimed at assessing whether developer claims hold up in use.

The second element is structured routes for sharing practitioner experience. Formal evaluation, including testbed evaluation, can establish whether a product works against its stated objectives, and is essential for that purpose. It is less well-suited to capturing the granular ways in which tools are used, modified, or worked around in real classrooms, and the conditions under which use turns out to be effective or ineffective in particular settings. A structured platform for practitioners to share experience of tools in real settings, closer in design to a peer review mechanism than a consumer review site, would complement formal methods rather than replace them. This type of user-generated evidence should feed into the Bayesian picture alongside other inputs, with appropriate weight given to its limitations.

Principle 8: Make evidence accessible to its intended audiences

Evidence about 'AI in education' must ultimately reach teachers, school leaders, and parents, not only researchers and policymakers. This requires active translation: from research findings into practical guidance, from academic language into plain English, and from lengthy reports into formats that policymakers and practitioners can use.

The question of intended audience should be settled before any evidence product is drafted, because it determines how the background is handled, what citations are needed, and how specific the guidance can be.

Glossary

Bayesian probability

A statistical approach where probabilities represent degree of belief in different competing proposals, and these beliefs are continually updated by combining existing knowledge with new evidence as it accumulates to produce an updated conclusion.

Cognitive offloading

The reliance on external tools to do thinking that a person would otherwise do themselves.

Context-mechanism-outcome lens

Asking not just whether something works, but for whom, in what circumstances and through what underlying process.

Inferential risk

The risk of drawing the wrong conclusion from evidence, and the fact that the costs of getting it wrong in one direction (for example, adopting something harmful) may differ from the costs of getting it wrong in the other (for example, missing something helpful). Definition should say specific strength of trial is it sheds light on causes - what effects what.

Precautionary principle

The precautionary principle is an approach to policymaking when there is scientific uncertainty about the nature and extent of a risk of harm to health but good reason to expect, and some evidence for, reduction in the risk of harm from reasonable action compared to taking no action.

Randomised controlled trial

A research study in which participants are randomly allocated to two or more groups, typically with one group receiving an intervention and another serving as a comparison group. Outcomes are then compared to determine whether the intervention has an effect. Random allocation helps ensure that differences between the groups are most likely attributable to the intervention, allowing RCTs to provide strong evidence about cause and effect.

Summative evaluation

The assessment of the impact of a technology after it has been deployed.

Bibliography

Hamelink, C. J. *The Technology Gamble: Informatics and Public Policy: A Study of Technology Choice*. (Ablex Pub. Corp., 1988).

Collingridge, David. *The Social Control of Technology*. (Pinter, 1980).

Stegenga, J. Fast Science. *Br. J. Philos. Sci.* https://doi.org/10.1086/729617 (2024) doi:10.1086/729617.

Stegenga, J. Down with the Hierarchies. *Topoi* **33**, 313–322 (2014).

Education Endowment Foundation. Teaching and Learning Toolkit. Teaching and Learning Toolkit.

Fesler, L., Martinez Clays, J., Agnew, C. & Loeb, S. *The Evidence Base on AI in K-12: A 2026 Review*. (2026).

Samson, R. & Pothong, K. *A Learning Curve? A Landscape Review of AI and Education in the UK*. (2025).

Edovald, T. & Nevill, C. Working Out What Works: The Case of the Education Endowment Foundation in England. *ECNU Review of Education* **4**, 46–64 (2021).

Huntington, B. *et al. Understanding Quality Characteristics of EdTech Interventions and Implementation for Disadvantaged Pupils*. (2025) doi:10.53832/opendeved.1170.

Orben, A. & Matias, J. N. Fixing the science of digital technology harms. *Science (1979).* **388**, 152–155 (2025).

Department for Education

ISBN: 978-1-83870-832-0

For any enquiries regarding this publication, contact www.gov.uk/contact-dfe.

This document is available for download at www.gov.uk/government/publications.