

Piloting the EdTech Evidence Board

**Report on key findings from phase one
and phase two**

September 2026

**Authors: Katy Chedzey, Sam Evans, Anna
Lindroos Cermakova**

Contents

Executive summary	3
Reframing the efficacy question	3
Key insights	4
Conclusions	7
Background and context	9
Overview of phase one	11
Aims and approach for phase one	11
Stakeholder engagement	11
Research	12
Key learning from phase one	13
The EdTech Evidence Board evaluation ‘core’ criteria	24
Challenges and opportunities	25
Proposal for phase two	27
Overview of phase two	30
Aims and approach for phase two	30
Advisory groups and steering groups	30
EdTech recruitment, selection and onboarding	31
Support for EdTech providers	33
Submission of portfolios	34
Portfolio review	34
The appointed board	36
Pilot outcomes	37
Key learning from phase two	38
Recommendations	51
What do these findings mean for the EdTech sector?	51
What do these findings mean for schools and colleges?	53
What do these findings mean for the system?	54
What does this mean for evidence assessment mechanisms, such as the EdTech Evidence Board?	56
Conclusion	60
Acknowledgements	61
The appointed board	61
Contributing individuals and organisations	62
Appendix A: EdTech Evidence Board evaluation criteria (including subcriteria)	63

Executive summary

Adoption of education technology (EdTech) continues to grow in schools and colleges in England yet effective and consistent use of technology remains uneven¹. Educators face a persistent challenge when making decisions about which products might most effectively meet their needs: while many EdTech providers make explicit claims about their product's ability to improve teaching and learning, there remains no consistent, independent approach to understanding the quality of the evidence underpinning those claims. Simultaneously, many EdTech providers are investing considerable time and resources into gathering evidence about their products, but often lack clear frameworks for demonstrating how that evidence supports claims about educational effectiveness.

The EdTech Evidence Board (EEB), funded by the Department for Education and delivered by the Chartered College of Teaching, was established to address these challenges. Across two phases, the project sought to develop and test an independent approach to evaluating the evidence for EdTech products while exploring how greater transparency and clearly defined standards for EdTech evidence might help to strengthen the evidence base for EdTech and ultimately enable better-informed decision-making across the education system.

Phase one ran from December 2024 to April 2025 and focused on developing a robust evidence evaluation framework through research, stakeholder engagement and co-design with educators, researchers and the EdTech sector. Phase two, running from April 2025 to April 2026, involved a small-scale pilot of the review process to test the criteria and the evaluation approach. Alongside the pilot, the project also developed practical resources designed to strengthen evidence capability across the education system.

Reframing the efficacy question

Traditional notions of EdTech efficacy tend to focus on demonstrating a product's ability to achieve its desired outcomes under highly controlled conditions. Recognising the limited application of such evidence to real world application of EdTech, the EdTech Evidence Board deliberately adopted a broader conceptualisation of efficacy, examining how a *range* of evidence might be able to provide a more holistic indication of a product's potential.

¹ Bolibruchova M, Hugill J, Rayman D, Yitbarek E. (2026) Assessment of the education technology market in England (Research Report). Department for Education. Available at: https://assets.publishing.service.gov.uk/media/6a3bcccf4c7605ab56723915/Assessment_of_the_education_technology_market_in_England_-_June_2026.pdf

Alongside traditional efficacy studies, this expanded approach to efficacy enabled the EdTech Evidence Board to look at theoretical evidence as well as evidence that examined the use and effectiveness of EdTech products in real world contexts. Whilst this is a step away from the more typical stance, the fundamental question being asked by the EdTech Evidence Board remains a question of efficacy: Is this product *capable* of working effectively and does it have the *potential* to achieve its intended outcomes?

Key insights

Recognising the potential of independent evidence review

Early stakeholder engagement showed that the concept of the EdTech Evidence Board was broadly well supported by both the education and EdTech sectors. One of the primary aims of the pilot was to determine whether a consistent, independent approach to evaluating EdTech evidence could be developed and applied in practice. The findings demonstrate that this is both achievable and valuable.

The evaluation criteria developed during phase one proved to be conceptually robust and practically workable across a diverse range of products and evidence portfolios. Independent reviewers were able to apply the criteria with a moderate degree of consistency, while moderation and governance arrangements supported transparent and defensible decision-making. The process demonstrated that evidence from multiple sources can be considered systematically without relying on a single hierarchy of evidence, allowing different forms of research and evidence to contribute to an overall understanding about product efficacy.

Participating organisations described the review process as rigorous, transparent and fair. Importantly, providers also valued the opportunity to receive structured feedback on their evidence, suggesting that independent review can provide developmental value alongside external assurance.

Together, these findings demonstrate that independent evidence assessment represents a credible foundation for strengthening confidence in EdTech evidence across the sector.

Takeaway: The pilot established that independent evaluation of EdTech evidence is both feasible and valued by stakeholders, indicating its potential as a mechanism for increasing confidence and transparency across the EdTech market.

Understanding the nature of EdTech evidence

Perhaps the most significant insight emerging from the pilot is that the principal evidence challenge facing the EdTech sector is more nuanced than is often assumed.

A commonly held perception is that EdTech providers do not generate or publish enough evidence. While evidence quality and availability undoubtedly vary across the market, the findings from this pilot suggest that many organisations are already undertaking substantial evidence activity. EdTech providers participating in the pilot frequently presented a broad range of evidence for consideration including implementation data, user feedback, case studies, internal evaluations and, in some cases, independent research.

However, relatively few portfolios were able to provide sufficient evidence to demonstrate a coherent account of efficacy that clearly connected educational need, product design, implementation, and outcomes - reflecting the holistic picture of efficacy set out within the EdTech Evidence Board's evaluation criteria. In many cases, evidence existed as fragmented, inconsistently communicated or insufficiently aligned with the claims being made about product effectiveness.

This has important implications for future practice. The findings suggest that strengthening evidence-informed practice in EdTech is not solely a question of encouraging organisations to generate more evidence. It also requires greater attention to evidence strategy, theory of change, claim discipline and the effective communication of evidence.

Rather than revealing an absence of evidence, the pilot highlights the importance of helping EdTech providers translate evidence activity into coherent evidence as practice.

Takeaway: The challenge is not simply about generating more evidence, but developing the capability to connect educational need, implementation and outcomes into credible accounts of efficacy.

Understanding evidence maturity across the EdTech sector

There appears to be considerable variation in evidence maturity across the EdTech sector, with EdTech providers being at different stages of their evidence journey whilst active in the marketplace. Initially surfaced through stakeholder engagement activities undertaken in phase one, this finding was further substantiated through the pilot. While some participating organisations presented broad and credible evidence portfolios, many were still developing the capability, partnerships and internal processes required to generate evidence meaning too few were able to meet the high expectations established through the EdTech Evidence Board evaluation criteria.

This finding should not be interpreted as a reason to reduce evidentiary expectations. Instead, it highlights the importance of creating clearer progression routes towards those expectations. Throughout the pilot, providers consistently reported that participation helped them identify evidence gaps, formalise approaches that had previously been implicit and strengthen internal understanding of evidence. For many providers, the review process functioned as a developmental exercise as much as an assessment.

This represents one of the most important findings of the project. While the EdTech Evidence Board was primarily conceived as a mechanism for independently reviewing evidence, the pilot demonstrated that the review process itself can strengthen evidence capability by helping organisations better understand how evidence should be planned, generated and communicated.

Supporting progression while maintaining rigorous standards may therefore be just as important as the assessment process itself.

Takeaway: Maintaining high evidentiary standards remains essential, but enabling organisations to build towards those standards is equally important if evidence quality is to improve across the sector.

Prioritising baseline compliance

Alongside its focus on reviewing evidence of efficacy, the pilot also examined a set of foundational prerequisites for products - including accessibility, data protection, cybersecurity, safety and interoperability - which revealed considerable variation in EdTech providers' ability to evidence compliance consistently across these foundational expectations. In several cases, providers believed they met baseline requirements but were unable to demonstrate this through consistent documentation or supporting evidence. Although these prerequisites do not provide evidence of efficacy, they represent essential foundations for trust and confidence.

At a system level, the pilot highlights the importance of providing greater clarity about what is expected of EdTech providers and helping them to better understand, evidence and communicate compliance with baseline standards more effectively.

Takeaway: EdTech providers may benefit from greater clarity around how to meet basic expectations and ensure baseline compliance with data, accessibility, safety and cybersecurity standards.

Understanding how evidence supports adoption

Throughout both phases of the project, educators consistently expressed strong support for greater transparency around EdTech evidence. However, they also highlighted that

evidence alone is rarely sufficient to support decisions about procurement, adoption and implementation.

Educators emphasised the importance of evidence being presented in ways that are accessible, meaningful and relevant to their own contexts. They valued concise summaries, implementation insights, contextual information and opportunities to learn from the experiences of other schools alongside formal research evidence.

The pilot also highlighted wider system factors that influence whether evidence is ultimately used in practice. Variability in evidence capability, limited collaboration between providers and researchers, challenges associated with generating evidence in authentic school contexts, and inconsistencies in how evidence is communicated and interpreted all affect the extent to which robust evidence can inform adoption.

Taken together, these findings suggest that improving evidence-informed adoption requires attention not only to evidence assessment, but also to the wider conditions that enable evidence to be generated, communicated, interpreted and applied effectively as a part of trust-related infrastructure.

Takeaway: Strengthening evidence-informed practice around EdTech requires coordinated action across EdTech providers, educators, researchers and the wider education system.

Conclusions

The EdTech Evidence Board was established to address a longstanding challenge within education: how to increase confidence in the evidence underpinning educational technology and support more informed decision-making across schools and colleges.

The pilot has demonstrated that independent evaluation of EdTech evidence is both feasible and valuable. It established a robust evaluation criteria, tested a transparent review process and showed that evidence from diverse sources can be assessed consistently and fairly. In doing so, it has laid important foundations for increasing confidence in the quality and credibility of evidence supporting EdTech products.

However, the pilot also revealed that the challenge facing the education system extends beyond the assessment of evidence itself. While many providers are already gathering substantial evidence about their products, there remains considerable variation in how effectively that evidence is planned, connected, communicated and used to support claims about educational effectiveness. Similarly, educators require evidence that is not only trustworthy, but also visible, accessible and meaningful within the realities of school need, discovery, adoption and implementation.

These findings suggest that improving evidence-informed adoption is unlikely to be achieved through evidence assessment alone. Lasting improvement depends upon strengthening the wider conditions that enable evidence to be generated, communicated, interpreted and applied effectively. This includes supporting stronger evidence capability across the EdTech sector, increasing the visibility and usability of evidence for educators, and encouraging greater collaboration between providers, researchers and schools.

The recommendations presented at the end of this report therefore build directly on the learning generated through the pilot. They retain the importance of rigorous, independent evidence review while recognising that assessment should form part of a broader system that supports continuous improvement in evidence quality and its use within education.

The work undertaken through the EdTech Evidence Board represents an important first step towards that ambition. By establishing common expectations for evidence, strengthening understanding of effective evidentiary practice and highlighting the conditions required for evidence to influence educational decision-making, the pilot provides a strong foundation for future work to support more evidence-informed adoption of educational technology across England's schools.

Background and context

Funded by the Department for Education and delivered by the Chartered College of Teaching, the EdTech Evidence Board (EEB) was established in response to a well-documented and growing challenge in the English (and global) education system: the proliferation of educational technology products in schools and colleges, coupled with a significant absence of reliable, independent evidence about their effectiveness. The EdTech marketplace is extensive. A recent report published by the Department for Education suggests that in 2026, there are at least 1,123 EdTech companies based in the UK². With so many products on the market, it can be extremely challenging for educators to determine which products and tools are effective, which ones are evidence based and which ones will work in their specific contexts. The market through which schools navigate this landscape has long been characterised by unverified claims, unsubstantiated impact data, and a lack of transparency about the evidence underpinning products. So much so that in a 2023 report published by BESA, it was suggested that just 8% of educators trust the claims made by EdTech companies³. More recently, findings published by Teacher Tapp put this figure at around 6%⁴.

The EdTech Evidence Board aimed to address this problem by taking an evidence-based approach to evaluating the efficacy of EdTech products using a defined set of criteria. The intention was that this would have a positive impact on the system by enabling educators to make better informed decisions around EdTech, helping them to feel more confident in choosing products that will make a real difference to teaching and learning in their settings.

The project was delivered in two phases. Phase one focused on developing the criteria that would be used to evaluate the evidence for EdTech products to provide a picture of their efficacy for teaching and learning. Phase two ran focused on piloting the criteria and review process with a small number of EdTech products across a limited number of product categories.

This report provides an overview of key activities within each phase, alongside sharing learning and insights, as well as recommendations for future work in this space.

² Bolibruchova M, Hugill J, Rayman D, Yitbarek E. (2026) Assessment of the education technology market in England (Research Report). Department for Education. Available at: https://assets.publishing.service.gov.uk/media/6a3bccccf4c7605ab56723915/Assessment_of_the_education_technology_market_in_England_-_June_2026.pdf

³ BESA (2023). The State of Evidence in EdTech. British Educational Suppliers Association. Available at: <https://www.besa.org.uk/wp-content/uploads/2023/03/230327-The-State-of-Evidence-in-EdTech-2023.pdf>

⁴ Teacher Tapp (2026). Digital Barometer Report Two: Summer 2026. Teacher Tapp. Available at: <https://aoztwr48.sibpages.com/>

Further information about the project can be found on the Chartered College of Teaching website: <https://chartered.college/edtech-evidence-board-project/>

Overview of phase one

Aims and approach for phase one

Phase one ran from December 2024 through to March 2025 and focused on developing the criteria that would form the basis for which evidence for EdTech products would be reviewed against. A secondary aim was to develop a workable approach to piloting and delivering the EdTech Evidence Board (EEB) review process in phase two.

To achieve these aims the project was delivered in three interlinked workstreams:

- 1 A **stakeholder engagement** workstream, which sought to maximise opportunities to engage stakeholders with the project to secure buy-in and help shape the design of an EEB system that would be effective in meeting the needs of the system
- 2 A **research** workstream, established to ensure that the EEB approach would be underpinned by research and evidence
- 3 A **design** workstream which drew on learning from the other two workstreams to inform the design of the criteria and pilot approach.

The interconnected relationship between these workstreams enabled the adoption of a co-design approach to the EEB. Learning from the research and stakeholder engagement workstreams was fed into the design workstream to inform the development of the EEB model. Design outputs were then fed back into the stakeholder engagement and research workstreams to test assumptions and confirm, refine or adapt the model as required.

Stakeholder engagement

The EdTech market is a complex environment which includes a range of key stakeholder groups, including not only educators and EdTech companies, but also EdTech associations, research institutions, accelerators and incubators, policymakers and more⁵. Whilst the primary focus in phase one was to engage with educators and EdTech suppliers, we also sought to engage with a broader spectrum of stakeholders to ensure our approach was representative of a range of sectors within the EdTech ecosystem, including but not limited to:

- 1 EdTech companies
- 2 Providers of EdTech services
- 3 Educators, schools and colleges
- 4 Experts and researchers

⁵ Lindroos Cermakova, A. & Havinga, B. (2024). Evidence in EdTech: Who needs it and why? European Edtech Alliance. Bielefeld, Germany.

5 Policymakers, influential sector organisations and regulators.

Stakeholders were provided with a range of channels through which they could engage with the project including individual conversations, events (including BETT), an open call for evidence, and educator and EdTech surveys and focus groups.

Through these activities, we saw active engagement in phase one from 73 EdTech suppliers, over 250 educators and representatives from over 25 EdTech and/or education sector organisations.

Advisory group

During phase one an advisory group was established to provide expert insight and feedback on the findings and recommendations emerging from the different workstreams, including providing feedback on design principles and early drafts of the criteria.

Involvement was voluntary and advisory group members were invited based on the specific skills and expertise they could bring to the project and with thought to ensuring representation from a range of key stakeholder groups. EdTech companies were not directly represented in the advisory group as it was considered that this could lead to an unfair advantage for subsequent phases of the EEB, however representation was sought from a number of EdTech sector organisations to ensure EdTech suppliers' views were represented.

Research

Research activity included a series of rapid research reviews looking at definitions of EdTech, EdTech quality and evidence frameworks, and digital pedagogy respectively; desk-based research reviewing the nature of information relating to certification and evidence claims on a small sample of EdTech product websites; a semi-structured survey and semi-structured focus group discussions.

435 respondents participated in the survey. The first question in the survey invited respondents to self-identify as one of four respondent groups:

- 1 Educator with decision-making responsibilities (educator-buyer) (33%)
- 2 Educator (any other educator) (27%)
- 3 Representative of an EdTech company (20%)
- 4 Other (20%)

Respondents received adapted questions based on their response to this first question.

Educators in total represented 60% of the respondents in the sample, however it is important to note that there may be some bias in the results as the sample is skewed in favour of respondents from secondary school settings. Based on the survey distribution methodology, it is also likely that the sample is skewed toward tech optimistic educators, which needs to be taken into consideration when looking at the findings. Respondents did not always respond to all questions, hence the number of responses to the individual questions vary considerably.

A total of 86 individuals from EdTech companies responded to the survey. These EdTech companies represented a range of products including tools for focusing on assessment and testing, content creation and assistive technologies and SEND. The greatest proportion of respondents (41%) specialised in products focusing on curriculum and subject specific offerings aimed at the K-12 market (Reception to KS5). As with the educator group, not all respondents from EdTech companies answered all of the survey questions and it is important to note that there may be some selection bias; that is, companies who are already thinking about evidence and have evidence of their products may have been more likely to complete the survey in the first place.

A number of respondents noted an interest to participate in further research for the project and these individuals were invited to participate in focus group discussions. Six focus groups were held in total with participants being grouped according to which group they self-identified as in the survey:

- 1 Educator with decision-making responsibilities (educator-buyer) - 11 in total
- 2 Educator (any other educator) - 13 in total
- 3 Representative of an EdTech company - 6 in total

In each of these discussions, participants were invited to share their thoughts and views on a variety of questions presented to them around EdTech, the use of evidence in relation to EdTech, and general perceptions towards the EdTech Evidence Board.

Key learning from phase one

Widespread use, but no shared understanding of 'EdTech'

Stakeholder engagement research with educators sought to understand the extent to which EdTech is used in education settings, and in what ways. As we might expect, it would seem that the use of education technology is fairly widespread; 82% of all educators responding to the survey reported that either 'everybody' or 'most' in their setting are using EdTech. In terms of the frequency of use, 92% of respondents reported that EdTech is used on a daily basis. However, there is a question raised as to what exactly is being referred to under the umbrella term of 'EdTech'.

The rapid evidence review found that whilst there is no singular agreed definition of the term 'EdTech', there's a general consensus that it broadly refers to digital and other connected technologies used to conduct or support education. A particular point of discussion is whether or not the term 'EdTech' refers only to products specifically designed for education. A recent guidance document from the US Department of Education⁶ takes two definitions:

- 1 Technologies specifically designed for educational use
- 2 General technologies that are widely used within educational settings.

Establishing a consistent understanding of what we mean by the term 'EdTech' is of central importance in terms of understanding the scope of the EEB. As such, a working definition was developed for the purposes of the EEB pilot:

By EdTech, we mean digital tools, applications, platforms and systems that are designed for education, and have the potential to support/enhance teaching and learning. This can involve anything from interactive digital books to online courses to virtual reality simulations and AI-driven assessment and tutoring. EdTech seeks to offer technologically enhanced teaching and learning experiences in ways that more traditional tools cannot, which does not necessarily imply higher quality.

Within focus group discussions held during phase one, educators were asked about their favourite types of EdTech products - this revealed insights into educators' perceptions of what they understand by the term 'EdTech'. Whilst our working definition of EdTech focuses on products that have been specifically designed for education, educators use of the term tended to be broader, including system software, such as Microsoft and Google, and also hardware devices, such as iPads, possibly even more frequently than more targeted solutions which will likely be the type of EdTech considered by the EEB.

There is a need to recognise that the term EdTech will be used in ways that extend beyond the working definition developed for this pilot; this will be an important area for the EEB to explore in future given the proliferation of such products within UK education settings.

Decisions around EdTech use are multi-layered

Stakeholder engagement research with both educators and EdTech suppliers looked at perceptions around EdTech decision-making: who is it that makes decisions about EdTech and what are the factors that influence such decisions?

⁶ U.S. Department of Education, Office of Educational Technology (2025). Designing for Education with Artificial Intelligence: An Essential Guide for Developers. Washington, D.C.

One key consideration for making choices around EdTech appears to be the importance of compatibility with what already is in the school, as highlighted by one educator: “tech that we feel we can integrate with our system that talk to other pieces of technology”.

When asked about how they choose technology, few educators had complete agency over which EdTech tools they use, with the majority indicating that they can ‘choose from a selection of EdTech tools’. However, it is frequently the case that educators are limited to tools selected by leaders. This has implications for the EEB which needs to meet the needs of these different users, and recognise that there are often likely to be two levels of influence around EdTech decision-making; the first, by leaders or decision-makers who may be purchasing the tools; the second, by educators who may choose whether or not to use the tools that have been purchased.

EdTech suppliers may underestimate the importance of evidence

When surveyed, educators generally agreed on which factors are important when considering implementation of educational technology, with the two most important factors being ‘user friendliness’, ‘educators’ needs and wishes’, followed closely by ‘regulatory compliance’, ‘research evidence’, and ‘ethical data use’. For educators with decision-making responsibilities, the importance of research and evidence in buying decisions comes through as an influencing factor, however this isn’t always readily available as one respondent commented: “I don’t think the research base is very strong. I don’t often come across convincing research evidence linked to edtech tools”.

When EdTech respondents were asked about customer priorities for making decisions, their perceptions aligned with educators in relation to the importance of ‘user friendliness’ as being one of the most important, however EdTech respondents also positioned ‘cost’ alongside this. Interestingly, factors such as ‘educators’ needs and wishes’, and ‘compatibility with the existing technological ecosystem’ were not particularly high on the agenda for EdTech respondents, despite having been highlighted by educators both in the survey and also in focus groups discussions.

Critically, less than 50% of EdTech respondents felt that research and evidence about the solution’s effectiveness was considered to be very important for EdTech decision-making.

A similar pattern was seen in the desk-based research carried out during this phase which looked at how EdTechs were using their websites to share information with potential users and/or decision-makers about their products. Though only a small sample of websites were reviewed, around half of the websites we looked at made claims about their products which covered a variety of different areas but tended to fit into one of four categories:

- 1 Impact on student outcomes (e.g. 150% progress achieved in English, maths and science across the academic year)
- 2 Research-informed basis (e.g. “Scientifically-proven” “Research-backed”)
- 3 User opinions (e.g. “94% of teachers say [product] adequately challenges students based on individual proficiency levels.”)
- 4 Workload (e.g. “4000+ hours saved this week”)

However, in all but a few cases it was impossible to substantiate these claims as the research or evidence was not shared. A common thread in the websites studied was the sharing of case studies / use cases. Often these were across a range of educational settings and focused closely on how a product could be used, rather than why, and with any claims about impact once again, not substantiated.

In terms of implications for the EEB, this highlighted that there may be opportunities to better support or encourage EdTech suppliers to gather more, or a broader range of, evidence about their products, and to raise awareness of research and evidence as a key factor influencing decision-making.

The idea of an ‘EdTech Evidence Board’ is well supported

Stakeholder engagement research showed that the idea of the EEB was broadly well supported by all stakeholders. When educators were asked whether they would find a platform such as the EEB helpful, an overwhelming 96% responded that they would.

EdTech survey respondents were similarly positive about future involvement in an initiative like the EdTech Evidence Board. When asked whether they would be interested in their company being included in a ‘go-to marketplace showcasing EdTech solutions with research evidence of impact on teaching learning’, and willing to invest in obtaining the required evidence, 93% of those who answered the question indicated that they would.

In the focus group discussions, the types of evidence that would be presented to the EEB was perceived as extremely important. EdTech focus group participants suggested they needed “clarity, e.g. clear ‘approved’ guidance or templates... to understand what is expected from them”. Educators stated that evidence included in the EEB should cover multiple perspectives, including case studies that show successful implementation in a similar context, with as much contextual detail as possible, echoing the findings of the survey, which showed educators’ preferences for others’ views and recommendations. They tended to be less enthusiastic about academic research: “it’s so wordy and usually very, very long. I want something that’s a quick read that I can quickly scan.” This highlights a need for the EEB to be able to communicate about EdTech evidence in a way that is accessible to educators but still sufficiently detailed for them to be able to make informed decisions.

Another point that was echoed by both educators and EdTech companies, was the speed at which technology innovates and that gathering evidence takes time. One participant stated “by the time you’ve published your research, it’s no longer up to date”, emphasising the need for the EEB to be dynamic and cumulative.

Both educators and EdTech companies were keen to understand more about what the EEB would look like; they were keen to know what the criteria would be, which products would be included and who would be involved in decision-making. Whilst we were unable to provide details at that point in the project, this underlined a need for transparency and the importance of clearly communicating the requirements and expectations of the EEB - and about what ‘good’ evidence might look like generally. As one EdTech company focus group participant suggested, it might be helpful “if we as product owners are not having to start from scratch and we are able to take guidelines that have been approved”.

Existing EdTech frameworks do not meet the specific needs of the EEB

The EEB evaluation criteria will play a key role in articulating evidence expectations for the EEB. As part of the rapid evidence review it was important to examine whether any existing EdTech Evaluation frameworks might be of use. The review identified a number of frameworks currently in use, focusing on a range of areas. Some frameworks used for pedagogical certification focused on systematic evaluation of EdTech by practising teachers and/or experts, with scoring usually aligned with established learning science theories. Other frameworks looked at broader aspects including technological and operational functionality.

Currently there is no one framework or standard that aligns completely with the aims and scope of the EEB, however, existing frameworks provide a helpful indication of quality and effectiveness to inform the development of evidence-based criteria for the EEB moving forward. For example, in a 2023 review of EdTech quality frameworks and standards⁷, Foster et al. analyse a range of existing frameworks synthesising ‘four commonly agreed overarching quality characteristics of EdTech product design:

- 1 EdTech products should meet the needs of its users
- 2 EdTech products should enable and support digital pedagogy
- 3 EdTech products should develop digital competence for learners and teachers
- 4 EdTech products should adopt an evidence-informed approach’ (p5)

These may be relevant to the EEB when it comes to evaluating evidence and considering potential indicators of efficacy.

⁷ Foster, D., McLemore, C., Olszewski, B., Chaudhry, A., Cooper, E., Forcier, L. & Luckin, R. (2023). EdTech Quality Frameworks and Standards Review. Department of Education

A broader conceptualisation of ‘evidence’ may be useful

Alongside EdTech evaluation frameworks, there are also a number of evidence evaluation frameworks, which typically focus on ‘strength of evidence’. Some of these frameworks are used solely to evaluate EdTech evidence, whilst others (such as the ESSA standards of evidence⁸ and AERO standards of evidence⁹) apply more generally to education evidence. These frameworks typically adopt a hierarchical approach to evidence. Whilst hierarchical approaches may present as an appealing way to categorise and communicate evidence quality, they reflect a tendency to promote internal validity (that is, the extent to which changes can be attributed to the intervention) over external validity (the extent to which the findings can be generalised more widely). Furthermore, such hierarchical classifications might be considered to reflect a narrow view of ‘quality’, predetermined by research design, with little consideration of how appropriate that research design might be, the extent to which the evidence may be favourable, nor the value of the different types of evidence that might help to understand the conditions that might contribute to success.

Foster et al.’s 2023 review acknowledges the challenges of looking towards a single marker of ‘quality’ for EdTech evidence and recommends the use of evidence portfolios ‘that allow an accumulation of varying types of evidence, encouraging and enabling an individualised approach to evidence-informed practice and decision-making (p7).’¹⁰

The idea that the EEB should value a range and breadth of evidence was one that was supported by EdTech providers and educators alike. Educators indicated that they were interested in, and found a clear preference for, more qualitative and contextually sensitive evidence that could perhaps provide insights into use and implementation; in responding to our survey, educators stated that the most relevant sources of evidence are ‘Teacher’s reviews and other practitioner recommendations’ and ‘Case studies explaining the contextualised use of the product’. Educators also indicated an interest in ‘Published research studies by research institutions (e.g. universities)’ but typically felt that ‘Research/evidence produced internally by the EdTech company itself’ and ‘Product certifications/badges by established private issuers’ were less relevant, perhaps indicating a lack of trust in research (and certifications) produced by private companies. This may strengthen the argument for the existence of a board such as the EEB, that can provide an independent evaluation of evidence for a product - it also emphasises the need for the EEB to be delivered by a trusted and reputable organisation. One educator commented that they would like to see a “Variety of research methodologies to create a holistic understanding of effectiveness”. This reinforced the hypothesis that there may be benefits

⁸ <https://www.evidenceforessa.org/wp-content/uploads/2024/02/FINAL-Standards-and-Proc-02-12-24.pdf>

⁹ <https://www.edresearch.edu.au/other/agency-publications/evidence/standards-evidence>

¹⁰ Foster, D., McLemore, C., Olszewski, B., Chaudhry, A., Cooper, E., Forcier, L. & Luckin, R. (2023). EdTech Quality Frameworks and Standards Review. Department of Education

in considering a portfolio of evidence for a product and the need to recognise that there is value to be gained from a broad range of evidence types.

Whilst it has already been highlighted that EdTech respondents perceived that educators do not place significant value on evidence, it was pleasing to note that the majority of EdTech respondents indicated that their organisation had engaged with research and evidence when designing and developing their product: 89% stated that they felt they engaged with existing educational research when designing their product, whilst 81% collected evidence about the effectiveness of their product before bringing it to market. This paints a positive picture of how research and evidence is used in early stages of product development.

It would seem that there is potentially less focus on gathering robust research and evidence once a product is established in the market. When asked about the current evidence held by EdTech providers for their product, responses indicate an inclination towards 'user data' and 'customer reviews' over more robust evidence sources.

It should be noted here that responses showed an unusually high proportion of companies that have evidence from EdTech testbeds and randomised controlled trials (RCTs) which perhaps reinforces once again that these findings may be based on a somewhat skewed or atypical sample. This aligns with wider insights gathered through informal discussions with a number of EdTech sector organisations who had collected data on evidence gathering from the EdTech companies they work with. Feedback from these discussions actually suggests that EdTech companies' gathering of research and evidence may be significantly lower than even the level indicated here. This leads us to raise a question about how we might make a distinction between 'research and evidence' and 'data' and how we might encourage EdTech companies to bridge the gap between the two.

EdTech suppliers experience barriers to evidence gathering

It's important to consider why EdTech companies may not engage in evidence gathering. In addition to the perceived lack of demand from education settings, EdTech companies also indicated that there was a lack of demand from funders and investors, with only 14 of the EdTech company respondents to the survey stating that funders asked for evidence of impact, which again indicates potential to explore how the EEB might work to generate demand for better EdTech evidence from not only educators, but also funders.

EdTech survey respondents also noted a number of obstacles to obtaining evidence with cost of research and the lengthy nature of the process being the most frequent barriers.

Informal discussions held as part of an EdTech summit hosted by the Chartered College of Teaching in March 2025 also highlighted the importance of viewing evidence gathering as

a collaborative process, and the need for educators and education settings to play an active role in this process, particularly given that the effectiveness of a product will be determined not only by the product itself, but in how it is used in practice. Furthermore, EdTech summit discussions also highlighted the challenges and tensions associated with collaborative evidence gathering, including EdTech providers not having access to schools and colleges in order to gather evidence, but also the potential demands this places on schools in terms of workload and time. Recognising the importance of gathering evidence in real-life contexts where products can be tested in controlled conditions (e.g. under supervision of researchers), we inquired in our survey as to whether the educators think their school would be willing to trial products in their setting to generate evidence about their pedagogical value and the majority of educator respondents (77%) were happy to consider the option.

EdTech companies participating in focus group discussions also noted that these were important elements; however, they pointed out that collecting consistent data sets in schools was “messy”, with one participant giving an example of a project that started collecting data from 160 children but ended up with only 15 complete datasets.

In the survey, 41% of EdTech companies stated they would need support to find research partners and 30% said they would welcome training materials that explain the basics of research related to producing evidence of impact. This again reiterates the point that there needs to be a way for educators and EdTech companies to work together to gather meaningful evidence and data.

These findings perhaps suggest that there may currently be missed opportunities for collaborative research gathering between EdTech companies and schools, and raises a question as to whether the EEB might play a role in bringing these two groups together.

A final consideration around barriers came from focus group discussions which highlighted that smaller companies or start-ups could face particular difficulty when it comes to gathering evidence, putting them at a disadvantage in terms of what they might be able to present to the EEB. Discussions considered the nature of research and evidence and a tendency for research to move in cycles, from prototype research (user research & usability testing) which was seen in the survey results, towards ‘research as evaluation’ which is iterative and only then progressing to the ‘RCT or equivalent’ as evidence is gathered. One participant in the focus group discussions gave a particularly helpful explanation of this research cycle:

“Once you start a prototype research, meaning user research and usability testing that is more about the kind of implementation aspect...you are focused less on trying to identify whether or not you are yet achieving the specific outcomes in your theory of change, but just checking, based on the user research and the ways that

people are engaging with your products, the theory of change is holding up until that point. And once you've got a level of confidence both in that initial product design in the way that that product is being used, you start to move it more into what becomes kind of research as evaluation to various degrees of maturity, whether or not that is in the first instance, just surveys to try and understand an initial indicative take on whether or not you're saving time or having an effect on pupil outcomes. And at the far end of that spectrum, for very, very mature products that have been in the market for a very, very long time, having an RCT or equivalent research tool to evidence that impact. And I think it's so important when we think about evidence that you have the right amount of evidence for the right point in the product lifecycle.” - *EdTech focus group participant*

This highlights that the concept of ‘research’ or ‘evidence’ may simultaneously hold multiple meanings and that the EEB should perhaps be open to considering how it might recognise the evidence for EdTech products at each stage of this journey - for example, by offering multiple review pathways. Indeed, participants in the EdTech company focus groups expressed a concern that the EEB might require companies to submit evidence of impact before they might be ready or able to gather this data. Alongside this, cost and access to researchers was also recognised as a potential barrier; it was noted that “large-scale research projects are unaffordable for start-ups for a fairly long time” and a suggestion was made that it might be helpful to involve “other professionals who are able to co-ordinate and do the research.”

This once again raises questions about whether and how the EEB might help encourage or facilitate greater collaboration between EdTech companies, researchers and schools.

A range of indicators could be used to evaluate EdTech effectiveness

Based on the findings of the stakeholder engagement and research activities carried out in phase one, it became clear that the EEB should look to a range of indicators when it comes to evaluating evidence of EdTech effectiveness.

A conventional approach might expect to focus solely on demonstrably evidencing the impact of specific EdTech products on learners (or on particular learning outcomes). However, this type of evidence can be costly to generate, takes time, and relies on EdTech suppliers having access to schools and colleges, as well as access to research expertise. That is not to say that evidence of demonstrable impact should not be the ultimate aim, but a singular focus on this risks alienating a significant proportion of the EdTech sector who may - quite understandably - not yet hold this kind of evidence. Furthermore, educators were quite clear that this was not the only type of evidence that they would find of value and a more nuanced approach to evidencing effectiveness would be required.

Emerging insights from phase one highlighted some further themes that would provide an indication of how the EEB might focus its attention around EdTech evidence more broadly. For example, The 2023 review of EdTech quality frameworks and standards¹¹ highlighted:

- 1 that EdTech suppliers should **adopt an evidence-informed approach**, gathering research and evidence to evaluate their product in order to understand its potential benefits and iterate design;
- 2 the need for developers to ‘understand the existing educational ecosystem’ (p6) in order to ensure products **meet the needs of its users** and **align with curriculum goals**;
- 3 the importance of **using research and evidence to inform product design**, in particular drawing from research relating to the learning sciences research and pedagogical best practices.

In addition, stakeholder engagement research in phase one indicated that:

- Educators saw value in **feedback or reviews** from practitioners who were already using a product. EdTechs similarly referenced customer feedback and also showed an inclination towards **user data** as a potential evidence source - each of which may be able to provide an indication of whether a product might actually be able to be used effectively in practice.

A recurrent theme which was often implied, but not necessarily surfaced explicitly, relates to the concept of digital pedagogy and how teachers are able to deploy the product effectively in their teaching in ways which would ultimately contribute to successful outcomes for their learners. A rapid evidence review explored the concept of digital pedagogy in more detail. A targeted search in the SCOPUS database¹² for research journal articles published in English between 2020 and 2025 based on the search phrase “digital pedagogy” yielded 267 results. Examining in detail a snapshot of 12 articles published in 2025, the research assumptions included in these articles suggest that the term ‘digital pedagogy’ is generally used in two distinct ways. Firstly, as a broad interpretation, encompassing any teaching practice that incorporates digital tools. Secondly, a more specific and narrow usage, emphasising the unique affordances that digital tools offer—opportunities that would not be achievable in traditional, non-digital classrooms.

¹¹ Foster, D., McLemore, C., Olszewski, B., Chaudhry, A., Cooper, E., Forcier, L. & Luckin, R. (2023). EdTech Quality Frameworks and Standards Review. Department of Education

¹² <https://www.scopus.com/sources>

Wegerif and Major¹³ highlight the importance of the concept of ‘affordances’ in EdTech, i.e., ‘what a user actually does with a tool, and the horizon of possibilities around that event – all the things that the user could imagine doing with the tool and all the associations that it has for them.’ Conole and Dyke¹⁴ go even further in suggesting that new digital technology has general affordances that can potentially change the nature of education. From this follows that different technologies have the potential of offering new and different pedagogical affordances.

‘Or, put another way, ... new digital technologies can enable ways of teaching and learning that were either not possible or more difficult to implement before these digital technologies became available’ (p5)¹⁵

The concept of affordances is important in relation to the complexity of real-life classrooms that include multiple variables that need to be considered when assessing how a tool will, or could, be used.

There are implications for the EEB in terms of the nature of the evidence that might provide insights into the **pedagogical affordances** of EdTech products. In relation to research methodologies, this reiterates the necessity to value a range of evidence, and a shift from a hierarchical approach that positions quantitative experimental studies such as RCTs as the gold standard, towards appreciating the role of qualitative and mixed methods approaches which may be able to provide richer and more nuanced insights that enable educators - and EdTechs - to better understand how to unlock the potential of EdTech in real educational contexts.

What this means for the EEB evaluation framework

These findings highlight that there is perhaps no single ‘ideal’ type of evidence for demonstrating the efficacy of EdTech products. Rather, the ‘best’ type of evidence will depend on what it is that those seeking the evidence wish to understand. In relation to teaching and learning, what educators tend to want to know at the highest level is: does the product actually do what it sets out to do? We might call this ‘actual’ efficacy. Being able to measure ‘actual’ efficacy may be ideal, but can often be challenging, particularly when it comes to determining a causal effect in real world education settings - it is well established that education settings are complex environments, and there are often many

¹³ Wegerif, R. & Major, L. (2024). *The Theory of Educational Technology. Towards a Dialogic Foundation for Design*. Routledge.

¹⁴ Conole, G. & Dyke, M. (2004). What are the affordances of information and communication technologies? *ALT-j*, 12(2), 113-124.

¹⁵ Wegerif, R. & Major, L. (2024). *The Theory of Educational Technology. Towards a Dialogic Foundation for Design*. Routledge.

factors and variables that can make a difference to the experiences and outcomes for learners and their teachers.

The EdTech Evidence Board therefore needs to recognise that it can be helpful to look at how efficacy may be evidenced in a multitude of ways; not just the ‘actual’ efficacy of a product, but also the conceptual efficacy of a product (in theory, could it work?) and practical efficacy of a product (is there evidence that tells us that a product is able to be used in a way that means it has the potential to produce its intended effects in the right conditions, or if implemented in a particular way?).

Learning from phase one has helped to identify six themes that each provide a useful lens through which to explore EdTech efficacy:

1. **Approach to evidence:** Does research and evidence inform product evaluation and development?
2. **Educational need:** Is there evidence to suggest there is a sound educational need for this product?
3. **Pedagogical design:** Is the design of the product informed by evidence?
4. **User experiences:** Is there evidence from users that suggests the product is able to be used effectively?
5. **Pedagogical affordances:** Does research/evidence demonstrate that the product can enhance teaching and/or learning?
6. **Impact:** Is there evidence to demonstrate positive effects on identified outcomes for teachers and/or learners?

These themes would ultimately form the basis of the evaluation criteria against which evidence for EdTech products was to be reviewed in the phase two pilot.

The EdTech Evidence Board evaluation ‘core’ criteria

Initial design work led to the development of six core criteria, informed by the research and stakeholder engagement activities carried out in phase one. These criteria were shared with the advisory group who provided useful feedback to help refine the criteria and clarify key terms.

The first three criteria are focused more on what might be considered to reflect ideas around ‘theoretical’ or ‘conceptual’ efficacy for the product. The final three focus more on ‘practical’ or ‘actual’ efficacy, requiring evidence to be gathered for the products that can provide insights about the use and effects of the products in educational contexts.

Criterion 1: Research/evidence informs product evaluation and ongoing development

Criterion 2: Research/evidence indicates there is a sound educational need for this product

Criterion 3: The design of the product is informed by evidence

Criterion 4: Data from users indicates that the product is able to be utilised effectively in schools and/or colleges

Criterion 5: Research/evidence demonstrates that the product can enable or enhance teaching and/or learning

Criterion 6: Evidence for the product demonstrates positive effects on identified outcomes for learners and/or teachers

Each criterion is exemplified by a series of subcriteria which provide an indication of how that criterion may be met in practice; these subcriteria are broadly hierarchical, intending to reflect progression within an evidence-gathering journey for a product.

Prior to commencing the pilot, the project team were able to test the criteria using sample portfolios. The criteria were iterated based on this testing however the general content of the criteria changed little during this iteration, which primarily involved reordering some of the criteria and minor wording changes where it was felt greater clarity was required - these changes were viewed positively when shared with the advisory group and other stakeholders for feedback. The full version of the criteria (inc. subcriteria) that were to be tested in phase two can be found in Appendix A.

Challenges and opportunities

Engaging stakeholders throughout phase one of this project highlighted some particular challenges that would need to be given careful consideration in the design, development and rollout of the EEB in phase two and beyond.

Challenge 1: Market readiness

There appears to be a gap between where EdTech companies are currently in terms of evidence-gathering for their products, and what the education sector might ideally want the evidence-base to look like. Additionally, the utilisation of evidence to inform decision-making and use of EdTech in education settings is perhaps also not as widespread or robust as we might hope for it to be. Even at a basic, conceptual level, there is not a widely accepted understanding of what we actually mean by the term 'evidence'. There is

opportunity here for the EEB to help to bridge these gaps by building better understanding around the gathering and utilisation of EdTech evidence within both the EdTech and education sectors. This highlighted a potential opportunity for phase two to consider how the EEB might facilitate the development of criticality and expertise in EdTech research, as well as promoting opportunities for collaboration between EdTech companies, researchers and education settings, helping to ensure that there is recognition that the effectiveness and success of an EdTech product is a shared responsibility which relies on effective partnerships between EdTech companies and those working in schools - as such there are clear benefits in both parties engaging in evidence-gathering to inform product evaluation.

Challenge 2: The need for the EEB to apply caution

A second challenge which had been a significant focus for the advisory group centred around the potential risks that EdTech might pose and how the EEB might effectively manage these.

In relation to submissions of evidence, there was recognition that evidence portfolios would potentially present a somewhat biased picture of a product given that it would be unlikely that EdTech suppliers would present evidence that portrayed negative or neutral effects. This was seen as being a particular issue if EdTechs were conducting research themselves about their products, and whilst it was recognised that externally commissioned evidence might provide a more balanced picture, there was still the chance that external researchers might still feel under pressure to present findings in a more positive light if they were funded by the company whose product was being evaluated.

To an extent, this challenge could be addressed in the pilot where robust assessment processes would ensure that those reviewing the evidence gave adequate consideration to research methodology and the corresponding validity of evidence claims. However, this also highlighted that there were ways in which the EEB could act to further mitigate these kinds of risks - for example by asking EdTechs to build risk mitigations into a logic model or theory of change for their products; requiring EdTechs to gather multiple sources of evidence relating to the effectiveness of their products in order to ensure claims were well substantiated; and supporting educators to adopt a critical approach to EdTech research when evaluating products' suitability for their context. Whilst it is beyond the scope of the pilot, there may be opportunity in future to consider how the EEB might draw from a broader range of sources - including evidence gathered by impartial organisations and evidence that has been gathered directly from education settings.

A second area of caution relates to the need to ensure that any EdTech products recognised by the EEB meet basic safety and legal requirements. Theoretically, the role of the EEB is solely to evaluate evidence of EdTech effectiveness; a thorough evaluation of

safety and legal compliance would be beyond the scope of the EEB, require substantial specialist expertise, and add both cost and complexity to the process. Nonetheless, this was clearly a matter that the EEB would need to address. With this in mind, a set of prerequisite criteria were developed which aimed to address this concern:

Prerequisite criteria

- i. The product meets basic accessibility requirements for the UK
- ii. The product complies with UK data protection regulations
- iii. The product complies with UK safety regulations
- iv. Proactive steps are taken to manage and minimise cyber security risks
- v. Steps are taken to support integration and compatibility with devices and systems used in UK education settings
- vi. The product is suitable for consideration by the EdTech Evidence Board

Developed in consultation with policymakers and the advisory group, these prerequisite criteria would be expected to be met before any product could be reviewed by the EEB.

Proposal for phase two

Learning from phase one highlights that there exists a clear and shared appetite across both educators and EdTech suppliers to engage more meaningfully with evidence, whilst simultaneously surfacing system-level challenges in accessing, interpreting and generating it in practice. Educators reported that evidence is valued but often difficult to find or use, while EdTech suppliers identified cost, complexity and capability as key barriers to producing robust evidence. This established a working hypothesis that we sought to test within phase two: That there is both demand for and interest in evidence-informed approaches across the system, but that existing structures, incentives and capabilities limit the extent to which this is realised in practice.

This informed a revised proposal for piloting an EdTech Evidence Board model in phase two that would aim to:

- a. encourage evidence generation and evidence utilisation in relation to educational technology
- b. provide a benchmark for evidence quality that is relevant, and of value, to the breadth of EdTech products within the marketplace, recognising that products and

companies may be at different stages in terms of product life-cycle and evidence maturity

- c. provide meaningful outputs and support educators, schools and colleges to engage critically with EdTech evidence and make more informed decisions about use of EdTech to support teaching and learning.

Instead of a single assessment pathway, the revised model would establish two separate review pathways for EdTech products with the option for EdTech suppliers to select the pathway most suitable for their product based on current levels of evidence maturity.

An **initial review pathway**, suitable for early-stage products with a more limited evidence portfolio, would review evidence against the first three of the EEB core criteria, examining evidence which sought to demonstrate the theoretical or 'conceptual' efficacy of the product. Portfolios would be reviewed by a single, trained reviewer and the review would aim to provide an early indication of product potential whilst also recognising that suppliers may require more time to develop their product's evidence-base.

Alternatively, a **full review pathway** would provide a more comprehensive evaluation of evidence for more established products whose evidence-base was more mature. Portfolios submitted to the full review pathway would be reviewed by a panel of trained reviewers who would evaluate evidence against all six of the EEB core criteria to provide a more holistic indication of product efficacy.

All products would be subject to **verification checks** against the prerequisite criteria prior to being accepted into the relevant pathway.

The utilisation of two review pathways intended to make the EEB review more accessible and relevant to a wider range of products within the marketplace, including newer products and products for which the evidence base was less mature. The intention being that products submitted for the initial review pathway would then be well-positioned to progress to the full review pathway in the future. This subtle change shifts the role of the EEB from purely serving a summative quality assurance function, towards having a more formative function, through which assessment against the EEB core criteria can act as both a mirror and a ladder: A mirror, because it could provide a realistic reflection of where a product is in terms of evidence of effectiveness at a snapshot in time; and a ladder because it could also provide a clear indication of evidentiary gaps which could inform the focus for future evidence generation.

Building system capability and capacity around EdTech evidence

It was also clear that there would be benefit in exploring how the EEB might work to build capability and capacity around EdTech evidence in both the education and EdTech sectors. Alongside the pilot, a range of resources would be developed including:

1. An EdTech Evidence 'Critical Evaluation Tool' - the tool would be co-designed with educators and would focus on using evidence to support decision-making around EdTech adoption
2. An online course in 'Building EdTech Evidence' and logic model tool - these resources would be developed in collaboration with experts and would aim to support those working in EdTech to build their understanding of how to plan and gather evidence for their products
3. Piloting of a researcher-matching programme which would pair EdTech suppliers with researchers and give access to tools and resources to facilitate effective EdTech researcher partnerships.

Each of these would serve to stimulate the future pipeline into the EEB, helping to prepare EdTech suppliers to generate better evidence for their products, and giving educators the tools they need to engage more critically with the EEB's eventual outputs.

Overview of phase two

Aims and approach for phase two

Phase two ran from April 2025 through to April 2026 and focused on piloting the 'prerequisite' and 'core' criteria (collectively referred to as the 'EEB evaluation criteria') and EEB review process with a small number of EdTech products across a limited number of product categories. Alongside the pilot, a second strand of work involved the creation of a range of resources that sought to build expertise and capacity around EdTech evidence for both the education and EdTech sectors.

Throughout phase two, timely and intentional engagement activities have enabled key stakeholder groups to input into the development of these resources. Wider stakeholder engagement activities have helped to maintain sector awareness of the project and its associated outputs, building buy-in and helping to establish a strong pipeline into the board.

Evaluation activities carried out throughout phase two sought to understand the potential for the EEB to achieve what it sets out to, including whether the criteria and review process are workable, scalable and effective; and the extent to which the EEB might play a role in improving evidentiary standards across the system.

Advisory groups and steering groups

The advisory group who were involved with phase one of the project were asked to continue supporting the project in phase two. It was decided at the start of phase two that as well as holding full advisory group meetings, stakeholders would be split into four steering groups depending on their area of expertise:

1. An educator steering group
2. A researcher steering group
3. An accessibility and inclusion steering group
4. An EdTech sector steering group.

As in phase one, it would not have been appropriate for representatives from EdTech companies to be part of the advisory / steering group due to the potential commercial advantage this might bring. However, it was important for us to continue to seek insights from those working within the EdTech sector; an additional focus group was therefore set up for this purpose.

These groups were invaluable in terms of the support and expertise they were able to offer during this phase, helping us to sense-check some of our assumptions and consider how key design decisions might play out. Many of these groups also contributed to the co-design of the capacity and capability-building resources developed in phase two.

EdTech recruitment, selection and onboarding

Recruitment and selection

The pilot sought to test the EEB evaluation criteria and review process with a minimum of 12 products which were considered to be directly related to teaching and learning, and which were currently available to education settings in the UK. To narrow the focus further, and facilitate helpful comparisons between evidence portfolios, it was also decided that the pilot would focus on just two product themes/specialisms. These specialisms would not be selected in advance, but were identified based on the makeup and characteristics of the products that were put forward by submitting organisations for inclusion in the pilot.

An open recruitment process ran from July 2025 through to September 2025 via an expression of interest (EOI) form accessed from the Chartered College of Teaching website. Recruitment utilised multi-channel approaches to maximise reach.

A total of 96 EOIs were received from EdTech organisations proposing products for inclusion in the EEB pilot. Data collected from the EOI was designed to give an overview of the product preparedness for potential review, specifically indicating: the product type/specialism; availability in the marketplace; evidence maturity stage; types of evidence available; and suggested assessment pathway, amongst other indicative questions. An initial review of submissions discounted any that fell outside of scope (i.e. there was not a clear enough relationship to teaching and learning or the product was not currently available to education settings in the UK) following which 54 submissions were considered proceedable.

Due to the limited number of spaces available for the pilot review round, a scoring system rubric was used to evaluate the EOIs. The rubric scored the EOIs against specific selection criteria:

- Motivation and commitment
- Strength of evidence portfolio
- Relevance to product categories and policy focus
- Accuracy of information
- Potential to reduce teacher workload
- Assessment of risk if admitted to pilot.

The rubric also took into account the need to find a balance of products in terms of product representation - termed 'balancing factors' these looked at: product maturity, evidence maturity, target educational phases/settings and product specialisms/use cases.

At the end of this process, based on the strength of submissions and alignment with policy relevance, eight products were identified for the full review pathway across two product specialisms:

- Formative assessment, pupil practice and quizzing
- Accessibility and inclusion (with a focus on literacy)

Selection was based on a comparative judgement of the relative strength and relevance of submissions, whilst applying the balancing factors to ensure sufficient representation across the two specialisms.

Given the emerging and rapidly-developing nature of the evidence base around AI in EdTech, the delivery team identified a potential risk in admitting AI-based products into the full review pathway. However, this presented an opportunity to intentionally select products that utilised AI for the initial review pathway, enabling us to evaluate the applicability of the criteria to this type of product. The majority of AI products put forward for the pilot were at an early stage of product maturity and therefore more suited to the initial review pathway than the full review pathway. As a result, a further five products were selected for participation in the initial review pathway which all utilised generative AI in some way.

In total, 13 products were selected for the pilot, exceeding the target recruitment number of 12. The decision to include an additional product helped to mitigate risk of falling below the minimum number should one of the pilot participants later withdraw or not progress through to submission.

Onboarding

Once selected, organisations entering the pilot progressed through a structured onboarding process designed to confirm suitability, ensure the integrity of submitted information, and provide targeted support for the development of a robust evidence portfolio.

Following the completion of verification checks against the prerequisite criteria, participating suppliers were given access to the EdTech Evidence Hub - an online platform developed for the pilot which hosted guidance and resources to support EdTechs in preparing their evidence portfolios for review. Each participating organisation was matched with an account manager within the project delivery team who guided them through the onboarding process and was available to offer informal support as required.

All but one organisation progressed through the entire onboarding process. This organisation felt that the prerequisite and onboarding process had given them a better sense of the requirements of the assessment process and the capacity required to enable their participation, and they expressed that they were unable to reallocate the required resources at that time. This meant that the pilot proceeded with 12 products - four for the initial review pathway and eight for the full review pathway.

Support for EdTech providers

In addition to the support available via the online EdTech Evidence Hub, participants had two scheduled check-ins with their account manager to ensure they were on track and prepare for submission. Organisations were able to request additional check-ins or get in touch with their account manager to seek clarification or confirmation about any aspect of the submission process as required.

Recognising the relative early stage of evidence maturity for those on the initial review pathway, an intentional aspect of the pilot design was to test how the EEB might act as an intervention for EdTechs seeking to enhance their evidentiary practices. With this in mind, those participating in the initial review pathway received two opportunities for direct support for evidence generation:

1. Logic model creation

Phase one surfaced the importance of structural artefacts in high quality evidence practice - in particular, that logic models may play a vital role in helping EdTech providers to demonstrate efficacy of their products by facilitating evidence planning and product evaluation, hence their inclusion in the EEB core criteria and wider work in phase two to develop an online logic model tool. Initial review pathway participants were invited to attend a bespoke webinar addressing the concepts of logic models as a strategic articulation of evidence practice and how to develop a logic model that could be incorporated into their evidence portfolios for review. Logic model submissions included in final portfolios from initial review participants were demonstrably informed by strategies articulated during this intervention.

It should be noted that this webinar was also made freely available to participants on the full review pathway and wider stakeholders, including those who completed the EOI but were not accepted onto the pilot.

2. Researcher matching and support

Recognising the need for the initial review pathway to work with early-stage products in a conceptual evidence arena, initial review pathway participants were

matched with a researcher who would work with them to produce two pieces of independent evidence that would help frame their portfolios:

- a literature review that would ground their design decisions of the product in appropriate pedagogical theory and curriculum alignment
- a collaboratively developed roadmap of recommended future evidence that would set the EdTech organisation on a stable path in implementing capacity to generate further evidence to help demonstrate efficacy.

Additionally, whilst initial review pathway participants were given the same opportunities of ongoing support from their account manager and the EEB delivery team via both scheduled and ad-hoc check-in meetings, the appetite for these was understandably higher and therefore accessed more frequently. The delivery team were cognisant of the need for impartiality and fairness within the process, which meant avoiding being overly instructive or supportive in ways that would lead to an unfair advantage. Instead, they typically answered questions and provided clarification around expectations. Rather than advise participants directly, account managers took a coaching-style approach - this was not something that was intentionally offered as part of the support programme, but which emerged naturally within the conversations and appeared to be a helpful mechanism for building confidence and evidence-awareness for those participating.

Submission of portfolios

Evidence portfolios were submitted for review via the online EdTech Evidence Hub.

All portfolios were checked by the delivery team to ensure they were proceedable before being shared with reviewers.

Portfolio review

Reviewer recruitment, selection and training

Reviewers were recruited via an open recruitment process in which they could apply for a role as either a single reviewer (for the initial review pathway) and/or a panel reviewer (for the full review pathway). A total of 119 applications were received.

Applications were anonymised and scored against the person specification. Scoring was carried out separately by two members of the delivery team. Early selection focused on the panellist role where applicants were categorised into one of three roles: educator, researcher and expert/specialist. Once scoring was completed, the highest scoring applicants were then tentatively grouped to form multi-disciplinary panels consisting of:

- An educator
- A researcher
- An expert/specialist

These prospective panels were then matched with portfolios that were perceived to align most closely with their skillset (for example, panellists with a background in formative assessment may have been allocated to a product that also focuses on formative assessment). The next stage of the selection process focused on selecting single reviewers for the initial review pathway and the highest scoring applicants were selected.

Fourteen reviewers (12 panel reviewers and two single reviewers) were taken forward into the second stage of the recruitment process which required them to undertake a practical task designed to assess research evaluation skills. Two reviewers were withdrawn from the process at this stage and their places re-allocated to the next high-scoring applicants. All applicants completed the task to a good standard, and the recruitment process concluded with 12 panel reviewers and two single reviewers appointed for the pilot.

Reviewers were provided with access to an online training course, which included templates and guidance for how to carry out the review process, along with practical examples. Additionally, panel reviewers attended an online training session held prior to their first panel meeting and initial pathway reviewers attended an online moderation session held with another reviewer and the delivery team.

Initial review assessment

The four initial review submissions were allocated and shared between two reviewers who reviewed their allocated portfolios independently. For criteria 1-3, reviewers made a judgement against each subcriterion before making a judgement against the overall high-level criterion using the descriptors provided. For criteria 4-6, which were not intended to be fully reviewed as part of this review pathway, reviewers were able to give an indicative judgement at subcriterion level, but did not provide a judgement against the high-level criteria.

Reviewers were provided with written guidance and a series of evaluation rubrics which they were advised to use to support them in evaluating the evidence submitted. Rubrics covered three themes: logic models, reviewing published literature and reviewing research evidence. These rubrics were developed by the delivery team, with input from a range of research experts, including the researcher steering group. Their design aimed to take a non-hierarchical approach to research evaluation, recognising the contribution and value that different types of evidence, including both quantitative and qualitative evidence can make to the evidence base for a product.

Once reviewers had completed their review, they shared the draft outcome report template with the internal delivery team. The delivery team checked the reports were complete and then shared them with the other reviewer for a second review. The two reviewers then met with the delivery team at a moderation meeting to discuss their reports and rationale for their decisions. Reports were finalised by the delivery team following this meeting.

Full review assessment

The eight full review submissions were allocated and shared between 12 reviewers who sat across four review panels. Each panel included an educator, an individual with research expertise, and a subject specialist (with expertise in either formative assessment or literacy/accessibility, depending on the product specialism). Panel reviewers were asked to read through each submission and then evaluate items of evidence individually to determine their credibility - and the credibility of any claims made. Reviewers were again provided with written guidance and evaluation rubrics.

Panel review meetings were held in late January 2026 across two half-day sessions. Two panel meetings were held simultaneously, with panellists joining for a 30 minute pre-meeting training session prior to the meeting commencing. For this pilot, members of the delivery team took on the role of chair.

For each product, review panels looked at each criterion in turn, first making judgements against the subcriteria, and then the overarching high-level criterion. For each criterion, the three panellists agreed upon one of four judgements: 'not met', 'partially met', 'met' or 'met (high)', reflecting the extent to which they were satisfied that they had seen credible evidence to support their judgement. Descriptors were provided to support these judgements.

A template outcome report was completed by the chair during the meeting, under advisement from the panel. The report captured all criterion and subcriterion level judgments and feedback for each criteria based on the panel's discussion.

The appointed board

At the end of the review process, outcomes are reported to the appointed board for ratification. The board consists of seven members whose role is to provide governance and oversight of the review process, ensuring that standards are consistently upheld.

It was typically expected that appointed board members would be sector leaders with the role description asking for experience in governance, senior leadership, EdTech and education research, with a strong understanding of the wider educational landscape, national policies and frameworks, systemic factors and the complexity of factors that may

influence practice around EdTech. As with reviewer recruitment, individuals were invited to apply via an open application process. In total, 74 applications were received and initial scoring took place carried out separately against the person specification by two reviewers. At this stage 32 applicants were identified to have not met one or more of the 'essential' elements of the person specification and were withdrawn from the process.

Applicants were then ranked by score and a second review of applications was undertaken for the highest scoring applicants. This second review focused on examining the skillsets of applicants in further detail, in order to ensure that the appointed board would have a diverse skillset and breadth of expertise to bring to the role. Discussions were held with shortlisted applicants to verify their suitability prior to confirmation of their appointment. A list of appointed board members can be found on page 60 of this report.

Appointed board members were provided with access to written policies and guidance materials, and attended a half day online training session prior to attending a full day outcome review meeting in which they reviewed all 12 outcome reports from the two review pathways. At this meeting, the board considered emerging patterns and trends and discussed criterion thresholds (i.e. establishing a consistent minimum bar for each of the criteria to be judged as having been 'met').

Following the review meeting, the delivery team applied these emerging thresholds to undertake a further round of moderation to ensure consistent application of the EEB core criteria across all of the portfolios. Moderated outcomes were presented to the appointed board for ratification at a final online ratification meeting which marked the conclusion of the EEB review process.

Pilot outcomes

Following the completion of the pilot, a joint decision was taken by the delivery team, its appointed board, and the Department for Education not to confer formal awards at this stage. As a first implementation, the pilot was designed to test the robustness of the evaluation criteria, the consistency of the review process, and the comparability of judgements across a diverse set of products and maturity levels - whilst of course providing tangible value to the participating organisations.

While many of the participating organisations demonstrated promising and credible evidence-gathering practices, the overall distribution and consistency of evidence against the criteria were not yet sufficient to support fair, reliable and comparable awards across the cohort. Issuing awards under these conditions risked establishing benchmarks that were not yet stable or sector-representative. Instead, the pilot prioritised the generation of detailed, independent outcome reports for each participant as an indicator of performance against the criteria, alongside system-level learning to strengthen the framework,

Each participating organisation received a copy of their outcome report in advance of a dedicated 90-minute meeting with a representative of the delivery team. The reports presented an indicative assessment of their evidence portfolio against the criteria, designed to provide a transparent and structured view of how each organisation's evidence aligned with the expectations of the framework.

Key learning from phase two

Through the pilot and wider activities undertaken in phase two, we sought to understand:

1. Whether the EdTech Evidence Board evaluation criteria and review process are workable, scalable and effective
2. The extent to which the EdTech Evidence Board might play a useful role in improving evidentiary standards across the system.

Additionally, phase one of this project highlighted a clear and shared appetite across both educators and EdTech providers to engage more meaningfully with evidence, alongside significant challenges in accessing, interpreting and generating it in practice. Educators reported that evidence is valued but often difficult to find or use, while providers identified cost, complexity and capability as key barriers to producing robust evidence. This established a **working hypothesis** that we sought to test within this pilot: That there is both demand for and interest in evidence-informed approaches across the system, but that existing structures, incentives and capabilities limit the extent to which this is realised in practice.

Alongside sharing insights about the effectiveness of the criteria and EEB process, this section of the report aims to surface holistic themes that provide a deep understanding of how the EdTech sector is currently engaged with evidence in practice. Together, these insights are intended to help inform both future recommendations about the development of the EEB and wider system-level thinking, highlighting where current approaches are working, where challenges persist, and which evidence ground may be most fertile for potential targeting of support in order to make evidence practices within the system more consistent and rigorous.

The learning presented here draws on a range of data sources gathered across the duration of the pilot combining modest quantitative outcome data with qualitative evidence from multiple stakeholder perspectives. The primary data sources include the scoring data in the evaluations: criterion-level and subcriterion-level outcomes for all reviews.

The experiences of participating EdTechs were captured through two complementary methods: A structured survey, completed by nine respondents across both pathways – provides breadth across the cohort on questions of clarity, burden, preparation, and

perceived value; and four semi-structured interviews, conducted after submission but before companies received any outcome information, provide depth on motivation, process experience, and the developmental value of participation. Reviewer experience was captured through a survey completed by five reviewers, a focus group attended by two full review pathway reviewers, and a single interview with one initial review pathway reviewer. The reviewer survey and focus group were conducted after the review process was complete but before final outcomes were ratified.

A note on sample size and generalisability is necessary. The pilot was, by design, small: numbers are not large enough to support statistical generalisation, and the findings should be read as illustrative based on the patterns and dynamics observed with this specific cohort that may provide indicative observations, rather than definitive conclusions, about the EdTech sector more broadly. The qualitative data is, however, rich – the interviews and focus groups in particular generated candid, detailed, and analytically productive accounts that would not be available from quantitative data alone, and the convergence of findings across multiple independent sources strengthens confidence in the observations made.

The EEB evaluation criteria and review process work in practice

The pilot found that the EdTech Evidence Board core criteria broadly worked well. Evaluation data from participating companies and reviewers suggests that both groups found the criteria to be clear and appropriate overall, whilst narrative responses provided helpful insights about where specific criteria or subcriteria could be improved or might benefit from minor adjustments in language to ensure clarity in how they are interpreted. Through the review, moderation and ratification processes, evidentiary thresholds were established which helped to build a picture of the minimum expectations for each criterion to be met.

The review process itself generally worked as intended - on an operational level the panel review approach was broadly successful, and processes and procedures worked well. The process has highlighted a particular challenge relating to the length of time it takes to review an evidence portfolio; many of our reviewers reported that it took longer than the anticipated timeframe to review submissions prior to the panel meeting or moderation meetings - which were also relatively time-consuming. This not only has cost implications, but with a finite amount of time available, will limit capacity - and ultimately scalability - of the EEB if this is not adequately addressed in future.

Verification checks revealed variability against the prerequisite criteria

Ultimately, the purpose of the prerequisite criteria is to test a core premise: That claims about impact are only meaningful if products meet baseline expectations for safe, secure, and accessible use¹⁶.

On the surface, the prerequisites appear to be straightforward and for many of the educators and professionals involved in our advisory group or steering groups, there was understandable expectation that organisations participating in the EEB should meet the highest of standards in relation to these prerequisites. In practice, the performance of submissions against the prerequisite criteria was variable. Verification checks utilised a self-report process with responses subsequently verified and RAG rated by the delivery team. This process found there to be a notable gap between where the education sector expected EdTechs to be and how these products measured up in reality, with disparity in how suppliers both understand and evidence compliance and variability across the domains. Data protection and security (perhaps driven by regulatory pressure) demonstrated the highest levels of compliance, followed by technical reliability and cybersecurity. The weakest area of compliance concerned accessibility and online safety. This uneven maturity presents a systemic risk, most notable for equitable access.

It is beyond the scope of the EEB to enforce these expectations, but the criteria can provide a signal to both EdTech and education sectors about what should be expected, and clear goals to work towards. It is clear that baseline assurance is not yet embedded in the EdTech market and there is an urgent need to formalise how compliance can be expected and articulated in future, allowing for the foundations in which trustworthy impact evaluation can rest. By creating the platform for EdTech organisations to be transparent about the extent to which they are meeting these prerequisite criteria, the EEB can play a role in revealing gaps and ultimately raising standards.

Performance against the EEB core criteria was varied - and weaker than perhaps may have been anticipated

We saw notable variation in terms of both the quality and breadth of evidence that was submitted in EdTech product portfolios, which led to variations in how these portfolios performed against the EEB core criteria. Overall, the criteria set a high bar which posed a challenge for many of the product portfolios submitted within this cohort.

Across the two review pathways, three broad tiers of performance were discernible. The first tier demonstrated credible evidence engagement across all six criteria, including criteria 4-6 which require evidence demonstrating what actually happened when the

¹⁶ Refer to p26 for an overview of the prerequisite criteria.

product was used in schools. The second tier showed solid performance on criteria 1-3, which assesses the conceptual and design-level evidence base, but fell away sharply at criteria 4 and beyond. The third tier struggled across the majority of criteria, though often for structurally different reasons.

This variation may perhaps be reflective of the fact that the different products are at different stages in terms of both product life-cycle and evidence maturity. However, at criterion-level a number of patterns emerge which are worthy of further discussion.

In relation to **criterion 1** (research/evidence informs product evaluation and ongoing development) the majority of submissions demonstrated an intent to use evidence for product evaluation and many were able to provide at least some credible evidence that they were doing so, however none of the portfolios were able to meet the high-level subcriteria within this criterion which focus on how EdTechs use research and evidence to inform the iterative development of their product and how this iteration is subsequently evaluated.

In relation to **criterion 2** (research/evidence indicates there is a sound educational need for this product) the majority of product portfolios performed well, however in many instances, reviewers had to work hard to extract the educational need from the evidence submitted. In stronger portfolios, educational need was often articulated through the use of an effective logic model or theory of change which showed how the product aligned to published curriculum or educational goals. This articulated educational need was then further substantiated by wider research or evidence such as published literature, market/user research or research which had been gathered about the product itself. In weaker portfolios pedagogical need and purpose were often poorly communicated, and in instances where evidence was used to validate the need, it was not always clear why a tech-based solution would be any more suitable than a non-tech solution. This suggests that there may be potential in supporting suppliers to understand how to develop products, and talk about these products in ways that are focused on ensuring clarity about the potential benefits the specific EdTech product may afford.

When it came to **criterion 3** (the design of the product is informed by evidence) it was frequently the case that the product aligned with research and evidence but it was not clear about how this had actually informed product design. Few of the submissions were able to meet the more demanding subcriteria - where they did, they were supported by a strong logic model and accompanying narrative which showed how research and evidence had actively informed design decisions. Perhaps the most demanding subcriteria within criterion 3 looked at whether learnings generated from evidence gathered about the product itself were fed back into product design. None of our submissions met this subcriterion which reinforces the findings in relation to criterion 1 and suggests that whilst there is an intent to use research and evidence, there is a gap in how that evidence is

being utilised once generated, and a missed opportunity to use evidence formatively to inform iterative product design.

Criterion 4 (data from users indicates that the product is able to be utilised effectively in schools and/or colleges) marks a shift in the criteria as it moves from examining conceptual/theoretical indicators of efficacy towards more practical or actual indicators. Criterion 4 also marked the point at which the cohort's performance considerably diverged most sharply from the preceding criteria. Where portfolios performed poorly, this was typically for one of two reasons: some simply did not submit any credible evidence relating to user experience, whilst for others, the credibility of the evidence submitted was too weak to demonstrate the criterion as being met. Interestingly, and perhaps somewhat surprisingly, few of the portfolios included evidence drawn from analytics or product-generated data; user feedback was the most dominant type of evidence submitted for this criterion but panels often found limitations in how generalisable the findings were, often due to methodological weaknesses in the evidence submitted. A common limitation was the lack of evidence gathered from a broad range of users and contexts - only one of the portfolios met this subcriterion. Mirroring earlier points relating to evidence informing product iteration, panels also saw little evidence of how user feedback was used to make improvements - and where this was seen, improvements were typically more focused on making tweaks to improve user experience (e.g. fixing bugs) than to inform fundamental changes to product design. Only one example within the cohort - a product on the initial review pathway - was seen to be systematically gathering feedback from users to inform product design. The mechanisms for gathering this feedback were an intentional design feature, embedded into the product itself. Where other organisations were having to retrospectively think about how they might gather feedback data, evaluation had been considered at the design stage and was gathered (and used) routinely to inform improvements.

For **criterion 5** (research/evidence demonstrates that the product can enable or enhance teaching and/or learning) it was perhaps expected that we might see an opportunity for EdTechs to submit evidence drawn from qualitative or mixed methods approaches which are able to generate rich insights about product use. However, few of our cohort were able to provide adequate evidence to show that the product enables or enhances teaching and learning and can be incorporated effectively into teaching.

Whilst some portfolios included case studies for this criterion, these were often marketing-driven and lacked methodological transparency. Generally, qualitative evidence was somewhat under-represented in our cohort which suggests that there may be work to be done to raise awareness of the potential and value of qualitative evidence for revealing rich insights about a product's use and affordances.

It was in relation to **criterion 6** (evidence for the product demonstrates positive effects on identified outcomes for learners and/or teachers) that we saw the weakest performance overall. Whilst many submissions included evidence that aimed to support this criterion, factors such as weak methodology, recency of evidence, and generalisability of evidence weakened evidentiary claims. Few organisations were able to provide credible and meaningful evidence that their product achieves its intended outcomes and demonstrates positive effects for teachers and/or learners. This was not entirely unexpected as given the challenges of extrapolating the effects of any educational intervention in real classroom settings, it was anticipated that it was at least possible that relatively few of the portfolios would meet this criterion.

The requirement for breadth and quality

As part of the review process, reviewers evaluated the credibility of the evidence items submitted within each portfolio, using this to inform their judgements against the EEB core criteria. Analysis of reviewers' evidence judgements indicates that participating EdTechs performed better in relation to theoretical evidence (literature reviews), whilst the quality of evidence gathered about the products themselves was generally weaker or more mixed. This aligns with the stronger outcomes we have seen for criteria 1-3 noted above.

For the full review pathway submissions, mixed methods and experimental studies tended to achieve more positive judgments. These tended to be carried out by external, independent evaluators, though reviewers were still often cautious about the findings.

Weaker performing evidence types included case studies, analytics and product feedback, surveys and correlational studies. The majority of these appear to have been carried out internally by the submitting organisation and were often characterised by a lack of transparency about methodology, weak research design, or by over-claiming or over-generalising findings. This could perhaps be indicative of a gap in expertise and low levels of evidence literacy within these organisations - which may be somewhat understandable, but could provide clear areas of focus for targeted support and improvement. However, it is not necessarily the case that externally commissioned studies were always more credible; a number of items of evidence submitted which were carried out by external/independent researchers or evaluators were flagged as being of low credibility, which is a particular concern, perhaps suggesting that even when EdTech companies may be seeking external expertise, there is still a gap in terms of the quality of expertise available to them, and the ability of EdTechs themselves to determine quality of the research outputs they receive. This reinforces the need for the EEB to provide a more visible benchmark for evidence quality - and help build understanding around evidence quality and raise standards across the system.

The lack of qualitative evidence and missed opportunities to gather insights about a product's use and affordances was noted in a number of review panel meetings where reviewers made an observation that they felt submitters were 'rushing to the end goal' and attempting to gather impact evidence too early in their journey where other types of evidence would have potentially added greater value.

Another point to note regarding the evidence submitted relates to the age of the evidence submitted. Typically, internally gathered evidence tended to be more recent, however externally generated and independent evidence was more dated. This is exemplified most clearly in the six experimental studies submitted of which four were over nine years old, with the earliest published in 2011 and the two most recent being from 2019 and 2022 respectively. We might hypothesise that lack of more recent studies could be reflective of the significant cost associated with carrying out this kind of research, or that submitting organisations may perceive that once they have this research they do not need to repeat the process. However, given that five of the six experimental studies were carried out within a pre-pandemic education system, and likely evaluated significantly earlier iterations of these products, reviewers were rightly cautious in their evaluation of these potentially dated evidence items.

The nature of the evidence seen in the initial review submissions aligned with what we might have expected to see from organisations submitting portfolios for products that were perhaps at an earlier stage in terms of product (and evidence) maturity compared to the products participating in the full review pathway. However, there were strengths in these evidence portfolios that were sometimes missing from the full review pathway submissions. For example, evidence was understandably more recent, and there was a clearer focus on gathering evidence to build understanding and check assumptions that had been made about the product. This could be conceptualised more as looking for 'concept validation' than seeking 'proof of impact'. Organisations tended to demonstrate better discipline in terms of their claims, with claims being made tentatively, minimal over-generalisation and greater awareness of evidence limitations. Due to their strong theoretical basis, two of the submissions within the initial review pathway actually performed better against criteria 1-3 than the lowest performing submissions on the full review pathway.

The outcomes described above highlight a key requirement that underpins the EEB design: That both quality/credibility of evidence and breadth of evidence are important. The portfolios that performed best were typically able to present credible evidence for their product, however this alone was not sufficient; organisations are also required to demonstrate breadth across their evidence portfolio if they are to satisfy the expectations set out within the EEB core criteria, which have been intentionally designed to provide a holistic picture of a product's efficacy for teaching and learning.

The EEB has the potential to play a useful role in raising evidentiary standards

‘Evidence activity’ versus ‘evidence literacy’

Findings from the pilot lead us to consider the important distinction between ‘evidence activity’ and ‘evidence literacy’. For the purposes of this report, we have developed working definitions to capture what is meant by these phrases:

Evidence activity: Actions taken by organisations to generate, collect or present information about their product’s use and impact. Evidence activity is concerned with what is ‘done’ - visible outputs and processes associated with evidence, like producing user surveys. However, on its own, it does not guarantee that evidence produced is methodologically robust or meaningful for decision-making.

Evidence literacy: The ability to understand, interpret and apply principles about what high-quality evidence is and how it is generated and utilised in a way that informs product development, evaluation and decision-making. Including the ability to define clear and testable claims, select appropriate methods and understand the limitations of different forms of evidence. Evidence literacy is concerned with how well evidence is understood and used, enabling organisations to move beyond activity towards evidence that is coherent, credible and aligned to intended outcomes.

While all participating organisations were actively engaged in generating or presenting evidence, this ‘evidence activity’ did not always reflect a secure understanding of what constitutes robust, decision-useful evidence. In practice, this manifested most typically as misalignment between claims and supporting evidence, limited articulation of causal mechanisms, and difficulty in interpreting or contextualising findings. This observed theme, in so much as we are able to extrapolate, suggests that the EdTech sector does not lack in evidence effort or intent, but rather in shared frameworks and fluency. Evidence generation is occurring, but not always in ways that are cumulative, comparable or methodologically sound.

Market logic and EdTech behaviour

A key learning from the pilot is that EdTech organisations’ approaches to evidence are often shaped less by methodological standards and more by the practical realities of how products are developed, positioned and sold within the education market. Within this context of market logic, forms of evidence that may be considered most effective for commercial engagement - such as case studies, testimonials and narrative-driven examples of impact - were most readily available for many of our pilot organisations to include within their submitted portfolios. While these examples may draw on useful data

that could provide insight into pedagogical effectiveness, market logic means the interpretation and mobilisation of this data instead focuses on market decision-making processes.¹⁷

Lindroos Cermakova et al. explain that the divergence between education logic and market logic 'does not imply incompatibility. Both logics are legitimate and necessary, and their coexistence is a structural feature of the EdTech ecosystem.'¹⁸ It's important therefore that market logic evidence is not viewed as being a product evidence limitation; it reflects the wider market environment in which EdTech operates.

Nonetheless, this raises questions about the role of market logic-oriented evidence within the EEB. It again leads us to consider the need to revisit how we develop shared understanding about what we mean by the concept of 'evidence' and how the EEB may encourage greater demand for, and interpretation of, high-quality evidence across the system.

Claim discipline and structural artefacts

A central learning from the pilot is the critical role of both claim discipline and the importance of structural artefacts (by which we mean logic models, theories of change, etc.) in a) underpinning high-quality evidence practice - as assessed through the review process; and b) setting EdTech organisations on a strong path towards evidencing efficacy. While closely related, these elements serve distinct but complementary functions, and the pilot found that together, they are often strong indicators of an organisation's ability to engage effectively with evidence standards.

We have adopted the term 'claim discipline' to refer to the ability of an organisation to clearly define and bound what its product is intended to achieve, what evidence objectively says about its performance and then how closely related these two are to impact claims. This includes specifying intended outcomes, target users and the conditions under which impact is expected to occur. Within the pilot, organisations that demonstrated strong claim discipline were able to articulate focused, testable claims that provided a clear anchor for evidence generation and assessment. This made it significantly easier to evaluate the relevance, sufficiency and quality of their supporting evidence.

In contrast, where claim discipline was weak - characterised by broad, multiple and overly-generalised claims - evidence portfolios often became diffuse and harder to assess. In these cases, organisations frequently presented large volumes of data or activity that lacked a clear relationship to a defined outcome, limiting its overall evidential value. This

¹⁷ Lindroos Cermakova, A., Havinga, B. & Clary, A. (2026). Needs-based EdTech Evidence Mapping. European EdTech Alliance. Bielefeld. Germany

¹⁸ Lindroos Cermakova, A., Havinga, B. & Clary, A. (2026). Needs-based EdTech Evidence Mapping. European EdTech Alliance. Bielefeld. Germany (p27)

highlighted that clarity of claim is not a superficial requirement, but a foundational condition for meaningful evaluation.

Structural artefacts played a crucial role in organising and operationalising those claims. These artefacts provide a structured representation of how a product is expected to lead to its intended outcomes, including the mechanisms through which change occurs and the assumptions that underpin them. Within the pilot, their presence enabled a clearer line of sight between product design, implementation and evidence, supporting both reviewers, and organisations themselves, in understanding how different evidentiary practices fit together. These artefacts can act as the ‘north star’ that helps organisations navigate the journey towards demonstrating efficacy, providing a clear path to follow.

Weatherby et al.¹⁹ highlight the importance of logic models and the vital role they can play in supporting communication, learning, and product development. This idea is strongly reinforced through analysis of pilot outcomes which similarly identified the logic model as a foundational component of high-quality submissions. In terms of performance, logic model quality often cascaded across multiple criteria and acted as either a structural asset or a limiting factor.

Importantly, the value of having strong claim discipline and well-considered structural artefacts has been seen to extend beyond the immediate requirements of the pilot review process. For many of our participating organisations, engaging with these elements prompted greater internal clarity around product purpose, strengthened the coherence of messaging and provided a more structured foundation for the future evidence development. Evaluation findings highlighted that the encouragement to generate artefacts such as these was perceived as being an unexpected benefit of the EEB process. In several cases, participating organisations reported that the process of refining claims and developing supporting artefacts helped to align internal teams, surface implicit assumptions, reflect on their true and intended purpose and identify gaps in both their evidence and product strategy.

Typical barriers faced by EdTech providers

The pilot also surfaced a consistent set of barriers that limited organisations' ability to develop and demonstrate high evidentiary standards as part of this process. These challenges were not typically the result of lack of intent; rather we would suggest they reflected structural, capacity and system-wide challenges that impacted on ability to translate intent into practice. Surfacing these challenges was a consistent theme

¹⁹ Weatherby, K., Clark-Wilson, A., Cukurova, M. and Luckin, R., 2022. The importance of boundary objects in industry-academia collaborations to support evidencing the efficacy of educational technology. *TechTrends*, 66(5), pp.784-797.

throughout the pilot outcome meetings, especially when considering the feedback from the reviewers for suggested steps organisations could take to strengthen their evidence base.

While the specific context of each organisation varied, a number of common barriers emerged:

- **Resource and time constraints:** Evidence generation is frequently competing for attention with product development, delivery and commercial activity. Limited time and financial resources constrained the ability to plan and execute rigorous evaluation approaches.
- **Data access and quality:** EdTechs universally articulated ongoing difficulties in accessing reliable, relevant and sufficiently granular data. This is well understood when working in school contexts, where data systems are fragmented and evaluation requests compete for attention from time-poor teachers.
- **Lack of shared standards and frameworks:** The absence of widely understood and adopted benchmarks for evidentiary standards contributes to inconsistency in both the quality and presentation of evidence - or even what evidence means and what counts towards it. Working in isolation of each other, the language of evidence efficacy differs from organisation to organisation and this subjectivity is communicated on to educational consumers.
- **Limited methodological expertise:** The majority of participating organisations lacked in-house research or evaluation capability to design, commission or critically interpret robust evaluations. This was often the most notable cause of some of the practice deficiencies stated earlier in this section.

Taken together, these barriers may help explain why, despite strong motivation, many organisations may struggle to produce evidence that meets higher standards of rigour. They also highlight that improving evidentiary practice is not solely an EdTech sector challenge, but one that sits within a broader system of infrastructure and support.

Engaging with the EEB has the potential to be transformative for EdTechs

Feedback gathered from participating EdTechs as part of the pilot evaluation shows that they found the support mechanisms that were put in place to be beneficial: the logic model support and the researcher matching initiative particularly, were rated highly by those who used them. The informal support provided by a dedicated account manager emerged consistently across the interviews as a significant enabler, particularly for smaller companies navigating their first encounter with an evidence review framework, for whom the availability of a knowledgeable, non-evaluative point of contact appears to have been as important as the formal guidance they received as part of the process.

However, learning from the pilot highlights the wider potential for the EEB to play a transformative role, and that the EEB review process itself can act as a critical mechanism for surfacing precise strengths and gaps in evidentiary practice. Through engagement with the EEB process, product evidence portfolios were not only assessed against a published standard; submitting organisations were supported implicitly and explicitly to better understand the relationship between their product, their claims and the evidence required to substantiate them. A consistent theme across all pilot participants was that this resulted in increased clarity, clearer articulation of intended outcomes, stronger alignment between claims and evidence and a more developed understanding of what constitutes robust and decision-useful evidence. In many cases, organisations were able to identify specific gaps in their current approach and crucially begin to define what would be required to address them. For the majority of pilot participants, engagement with the EEB was initially motivated by the outcomes they would ultimately receive and the desire for recognition. Yet it became clear that in reality, it was the experience of the EEB process itself that led to powerful new realisations and insights. Many overestimated the quality of their portfolios and benefited from the cognitive dissonance or ‘aha moments’ that occurred when preparing their portfolios and receiving the outcomes of their reviews. This phenomenon was not unexpected - it is widely recognised that accreditation has the potential to shape and disrupt beliefs and practice, acting as a powerful catalyst for development.

The value of the pilot therefore extended beyond individual assessment outcomes. It has provided each participating organisation with a clearer baseline from which to develop, equipping them with the language, frameworks and insight needed to strengthen their future evidentiary practice. In this sense, the EEB process has demonstrated its potential not only as a mechanism for evaluating efficacy through evidence, but as a driver of capability or capacity building from aspirational evidence towards a more confident ability to demonstrate efficacy over time.

Summary

The phase two pilot has helped to validate our initial hypotheses: It is feasible to assess EdTech evidence against a structured framework (the EEB core criteria). There is both demand for, and interest in, evidence-informed approaches across the sector, but existing structures, incentives and capabilities limit the extent to which this is realised in practice.

Throughout this pilot we have seen that appetite for the EEB is high but there was often a ‘rush’ towards demonstrating measurable impact too soon, and to the detriment of other evidence that could provide useful and actionable insights in terms of product effectiveness. Across participating organisations, there was consistent evidence of high engagement and willingness to invest time and resources in evidence-generation. However, this was accompanied by variability in the quality, coherence, and methodological robustness of submitted evidence, corroborating many of the perceived

barriers that were identified in phase one. This was most clearly evident in the outcomes for the pilot reviews in which we saw a clear structural divide between organisations' ability to articulate intent with their ability to demonstrate impact in practice.

Substantiated through the pilot is the notion that when given structured frameworks, clear expectations, and the opportunity to engage in a rigorous but developmental process, participating EdTech providers were able to make meaningful progress in how they articulated, organised and reflected on their evidence.

The crux of the phase one hypothesis - that the EdTech sector is willing but not yet fully equipped to operate at higher evidentiary standards - has therefore been upheld. The pilot has offered further insight to indicate that there would be value in investing in a mechanism such as the EdTech Evidence Board that can:

- a) enable capacity and capability-building around EdTech evidence
- b) set clear evidentiary standards for EdTechs to work towards
- c) provide increased evidence-visibility to support EdTech adoption and decision-making in education settings.

Recommendations

What do these findings mean for the EdTech sector?

The findings of this pilot challenge a common assumption that the principal evidence challenge facing the EdTech sector is a lack of evidence generation. Across the pilot, participating organisations frequently demonstrated substantial evidence activity, however few were able to sufficiently demonstrate a clear rationale connecting educational need, product design, implementation, outcomes and evidence into a coherent account of efficacy. Furthermore, the pilot surfaced inconsistencies in how EdTech providers were able to demonstrate compliance against the foundational prerequisites (data and privacy, accessibility, cybersecurity, operability, and safety) that need to be addressed if schools are to have the confidence that products meet baseline expectations for safe, secure, and accessible use.

Learning from the pilot suggests clear actionable steps that EdTech providers could take to help strengthen practice in these areas:

1. **Ensure baseline compliance with data, accessibility, safety and cybersecurity standards.** These are the foundational requirements for any product to be used effectively in schools and play an important role in the decision-making process for EdTech procurement. Compliance needs to be an expectation, not an aspiration. EdTech providers must ensure they have a sound understanding of current expectations, including the online safety act, UK GDPR, the Data Protection Act 2018, Data Use and Access Act 2025 and, where relevant, guidelines such as the DfE's Generative AI product safety expectations²⁰, WC3 Web content accessibility guidelines²¹, and the ICO's Age appropriate design: A code of practice for online services²². EdTech providers should regularly audit practice against relevant guidelines and requirements and where gaps are identified, a clear and timebound plan should be developed and implemented to progress towards compliance at the earliest possible opportunity.
2. **Ensure products have a well-considered logic model and use this as the lynchpin for strengthening and demonstrating effectiveness.** A useful logic model should demonstrate how a product works to meet an identified educational need and help to drive evidence gathering. It should be viewed as a dynamic

²⁰ <https://www.gov.uk/government/publications/generative-ai-product-safety-expectations/generative-ai-product-safety-expectations>

²¹ <https://www.w3.org/TR/WCAG22/>

²² <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/code-standards/>

artefact and will evolve as providers gather evidence that helps to validate assumptions and inform product iterations.

3. **Establish a coherent evidence strategy that values a breadth of evidence and embeds evidence gathering and utilisation within product development processes.** Logic models can provide a foundational role in a product evidence strategy, but providers should be wary of rushing to demonstrate impact too soon. The EEB evaluation criteria provide a useful framework for auditing existing evidence in order to identify gaps and inform next steps. A strategy that seeks to gather a range of evidence may be most helpful when it comes to evidencing effectiveness. In particular, EdTech providers should consider how they gather evidence about how their product is able to be used effectively in terms of both user experiences and pedagogical affordances. This type of evidence can help to provide earlier indications of product effectiveness and has the potential to offer greater insight about where improvements could be leveraged to strengthen the impact of a product in the longer term. Consideration should also be given as to how evidence gathering and utilisation can be embedded systematically into internal processes so that it becomes part of routine practice, rather than being perceived as an add-on. Providers might consider how they could embed opportunities for gathering data for product evaluation directly into the design of their products in order to provide a reliable pipeline of evidence and maximise efficiency. Similarly, there should be a clear mechanism for funnelling insights from evidence into product development and logic model iterations.
4. **Commit to developing internal understanding around evidence and practice sound claim discipline.** Building evidence literacy - that is, the ability to understand, interpret and apply principles about what high-quality evidence is and how it is generated and utilised in a way that informs product development, evaluation and decision-making - is vital if providers want to build credibility and trust in product evidence claims. Alongside their product evidence strategies, EdTech providers should look for opportunities to strengthen shared understanding of effective evidentiary practice and support individuals and teams to develop specific expertise in areas relevant to their roles where this may offer further value.

There are a range of resources that have been developed as part of the pilot that are designed to support EdTech providers in taking these actions forward. These have been made freely available and can be accessed via the Chartered College of Teaching website²³.

²³ <https://chartered.college/edtech-evidence-board-project/edtech-resources/>

What do these findings mean for schools and colleges?

It was clear from our work in phase one that teachers and leaders working in schools and colleges would value greater evidence visibility around EdTech to help inform decisions about whether - and how - products might offer potential solutions and meet the needs of their contexts. Whilst the majority of work in phase two focused on piloting the EdTech Evidence Board approach with EdTech providers, educator input played a critical role in shaping key elements of the programme. Engagement through the educator steering group provided valuable insight into how educators interpret, interrogate and make decisions based on evidence in practice. Equally, whilst not a practical reality of the pilot, educators provided clear advice on the public-facing outcomes from our assessments - expressing the necessity for them to be accessible, meaningful, and usable within the realities of school life. This group also worked with us to codesign an EdTech Evidence Critical Evaluation Tool (known as 'EdTech Check') which aimed to provide a helpful structure for engaging with EdTech evidence for decision-making and procurement and has been well received by the sector.

The findings of the pilot highlight that a notable proportion of the EdTech sector may not currently be ready to provide the credible evidence that educators are looking for. The recommendations captured here therefore intend to offer suggestions as to how those working in schools and colleges might navigate this potential mismatch between expectations and reality.

- 1. Adopt a critical approach when evaluating product claims.** It's vital that educators ask questions of EdTech providers about the evidence base for their products, and scrutinise the claims that are being made: What evidence is this claim based on? Who were the participants? What methods were used? Is it likely that this claim could be biased or flawed in any way? The EdTech Check tool provides advice and prompts to guide this process and is freely available via the Chartered College of Teaching website²⁴.
- 2. Utilise a range of evidence to monitor and evaluate the effectiveness of EdTech products in your setting.** While schools and colleges should expect to be able to look to EdTech providers for evidence of product effectiveness, there is still a responsibility to gather contextual evidence to help teachers and leaders understand whether the products that are adopted in their setting are being effective in meeting their needs. The nature of this evidence will vary depending on the purpose of the product, and whilst student outcome data is often perceived as being most desirable, it is usually difficult to attribute to a specific intervention such as use of a particular product. Furthermore, it usually takes time for the impact of an

²⁴ <https://chartered.college/edtech-evidence-board-project/resources/>

initiative to be visible in outcome data and schools don't always have the luxury of time to wait to see if a product has made a difference. It can therefore often be more helpful to look at proxies or to gather evidence that might provide early indications of success. Additionally, there's value in gathering insights around implementation and pedagogical use in order to understand what works well and identify barriers or issues. It is helpful to plan evidence gathering for monitoring and evaluation as part of any product implementation strategy, and to continue to check on product effectiveness on an ongoing basis once a product is established.

- 3. Consider whether you may be in a position to collaborate with EdTech providers and researchers to strengthen the evidence base.** Many of the EdTech providers engaged in the pilot reported that access to education settings was a significant barrier to evidence gathering. Evidence generated from EdTech research in schools and colleges can provide valuable insights for the EdTech providers, the settings themselves, and helps contribute to the knowledge base of the wider system. Currently, insufficient evidence exists around product effectiveness which means that leaders are having to make decisions about products based on need alone, without having the confidence that the products they are choosing to adopt are actually capable of meeting that need. This poses a clear risk for learners that can only be mitigated by a shared commitment to gathering evidence so that in the longer term, decisions can be based on evidence as well as need. Of course, leaders should apply a degree of caution whenever it comes to carrying out research in their settings and should expect researchers to ensure that any research is carried out ethically and without the risk of harm to learners or teachers.

In relation to how educators might utilise a resource such as the EdTech Evidence Board in future, it was clear from our work with educators across both phases of this project that educators want to have access to a range of evidence insights alongside more practical information about products relating to areas such as cost, safety, accessibility, data protection and compatibility. Whilst many of these are present within the prerequisite criteria, they currently sit outside the scope of evidence evaluation undertaken by the EEB. Nonetheless, consideration should be given to how these might integrate into the outputs of a system such as the EEB in future, and in the meantime, educators can utilise the EEB evaluation criteria as a structure for thinking about both evidence and the prerequisites for adoption.

What do these findings mean for the system?

Learning from across the two phases of this project highlights that the EdTech market is shaped not only by the behaviour of EdTech providers, but also by the expectations and

incentives created by the wider educational system. Whilst educators consistently expressed a desire for trustworthy and accessible evidence during phase one, the pilot highlighted a more complex reality: the types of evidence that are most visible within the market are not always those best suited to understanding efficacy. This raises important questions about how evidence is valued and utilised across the wider system.

Improving educators' access to product evidence alone may not be a sufficient solution to ensuring more efficacious tech occupies the classroom; there is also a need to strengthen evidence literacy amongst those responsible for evaluating, selecting and implementing technology. The pilot highlights the extent to which procurement and adoption decisions shape provider behaviour. If schools, trusts, funders and policymakers place limited emphasis on interpretation of complex evidence when making decisions, there is perhaps little incentive for organisations to invest in the time, expertise and resources required to generate more robust evidence. Conversely, where expectations for evidence are clear, transparent and consistently applied, organisations have strong reasons to invest in evidentiary practice. Improving the quality of evidence available within the market is therefore not solely a provider challenge, but a system challenge that requires demand-side change alongside supply-side improvement.

Recommendation 1: Future system-level initiatives should seek to strengthen both the supply of credible evidence and the demand for its meaningful interpretation and use.

A further finding from the pilot concerns the relationship between evidence and trustworthiness. Verification checks identified notable variability against prerequisite criteria, particularly in relation to accessibility and online safety. Whilst these criteria were not designed to assess educational impact directly, they highlight an important principle: Evidence of efficacy is only meaningful when products can also demonstrate that they meet reasonable expectations for safe, equitable and responsible use. The findings suggest the wider system may benefit from providing greater clarity about what is expected from providers and how they might demonstrate compliance, placing greater emphasis on baseline assurance that products meet minimum expectations.

Recommendation 2: Future system-level initiatives should encourage greater trust in EdTech by establishing clear and consistent expectations for both evidence quality, and foundational prerequisites (such as accessibility, online safety, etc.).

The future success of evidentiary standards will depend as much on how evidence is consumed as how it is produced. Better evidence alone will not necessarily lead to better adoption and therefore greater effect in context. The wider system must also create the conditions in which evidence is expected, understood, trusted and acted upon.

This raises important questions for the future: How can the system establish clearer more consistent expectations about what constitutes credible evidence of effectiveness? And what conditions are required for evidence to become a meaningful driver of product development, procurement and educational decision-making?

What does this mean for evidence assessment mechanisms, such as the EdTech Evidence Board?

The EEB pilot demonstrates that there is a sound case for a future mechanism for evidence assessment, but that there is potential for this to function in a way that moves beyond evidence assessment alone, towards offering a more holistic function that enables capacity and capability-building around EdTech evidence, and helps to provide increased evidence-visibility to support EdTech adoption and decision-making in education settings.

With this in mind, the recommendations presented here aim to offer suggestions as to how these aims might be achieved.

Recommendation 1: Explore how the outputs from evidence assessment mechanisms might be published alongside more nuanced insights and practical information to better support EdTech procurement, adoption and implementation in schools and colleges.

Throughout the two phases of this project, educators have consistently told us that they would find value in a resource such as the EEB where they could access reliable information about the evidence-base for an EdTech product, but they also indicated a preference for such a resource to include broader information about EdTech products - such as information around value for money, operability etc. - which would sit outside of the EEB's current scope. While it is clear that this broader information does not form part of evidence assessment, there is an opportunity for a future system of evidence assessment and assurance to provide a space for curating a range of information in a way that is usable for busy school decision-makers.

When it comes to evidence visibility, evidence assessment mechanisms may need to be creative in how evidence is communicated to preserve nuance, whilst providing clear and actionable insights. Information should be published in a consistent format enabling educators to compare products and quickly identify gaps, in order to inform decision-making. There is a need for proactive activity to support educators to engage critically with evidence as part of this process, but there is also an opportunity to surface insights generated from the education sector directly. Future phases of an evidence assessment mechanism could explore how to leverage additional mechanisms for evidence generation and learning around EdTech. For example, it may seek to facilitate knowledge sharing and evidence generation in schools and colleges about specific EdTech products, or around

EdTech effectiveness more generally. In particular, we have seen from our work on EEB a clear lack of evidence relating to the pedagogical affordances and effective implementation of EdTech products and finding an efficient way to draw upon and amplify insights from schools and colleges would provide an opportunity to address this gap.

Recommendation 2: Further stage the review process within an evidence assessment mechanism to reflect the differing levels of evidence maturity within the EdTech market, and build a strong pipeline towards high evidentiary standards, such as those defined in the EEB evaluation criteria.

The pilot also surfaced insight about the relationships between evidence quality and evidence maturity. Whilst the review process demonstrated that robust assessment is feasible, the distribution of outcomes suggests that many organisations are still developing the capabilities required to consistently produce forms of evidence that the EEB would expect. This does not imply that standards should be lowered. Rather, it suggests that future approaches may need to further distinguish between how the process can support EdTech providers to develop stronger evidentiary practice and independently assess the outputs of that practice, when ready.

Insights from our work with EdTech providers participating in both the initial and full review pathways suggests that the gaps between current and desired evidentiary practice could be addressed by leveraging an evidence assessment process as a developmental mechanism. With this in mind, there may be benefit in recognising progression across different stages of evidence development whilst maintaining clear and consistent expectations around the quality, credibility and appropriate interpretation of evidence. This would enable organisations to demonstrate meaningful progress in evidentiary practice without diluting expectations around rigour or trustworthiness.

A verification process could be the gateway through which EdTech organisations are able to access an evidence assessment process. The process should not attempt to assess evidence relating to the prerequisite criteria but should be a platform through which EdTech providers can share information to demonstrate how their products meet these prerequisite criteria, and be the gateway for accessing capability building resources. Combining this with an initial self-evaluation against defined evaluation criteria could support EdTech providers to ascertain a picture of current evidentiary practice and help to surface gaps to inform future evidence strategy.

Introducing further staging in the assessment process could help to alleviate the challenges faced by EdTech providers in meeting the more challenging criteria. For example, a sound logic model has shown to be a key factor that underpins effective evidence generation and valid product claims, so this may be a useful first step in a staged approach, along with evidence that supports that there is a sound educational need for the

product. This could also offer a way of operationalising evidence assessment mechanisms to address recommendation 1 above: product information could be published on a digital dashboard, along with information gathered about the prerequisite criteria as part of the verification process. This would give schools and colleges access to key contextual information about each product and could then form the basis of a digital evidence portfolio that would capture all outputs from the process as products progress through a staged review pathway. Beyond this, EdTech organisations could continue to add to their digital portfolio, submitting individual evidence items for review as they are developed, which would also help to reduce the time-burden for reviewers. Such an approach would enable products to progress from one stage to another, receiving recognition and promoting ongoing engagement, helping to facilitate the development of a strong evidence culture over a sustained period of time. It would also provide a clear route to raising evidentiary standards across the system, enabling EdTech providers to engage meaningfully with the system in a way that is appropriate for their product in terms of both product lifecycle and evidence maturity. In such a model, capability-building becomes a core function rather than merely a supporting activity.

The pilot suggests that one of the most valuable roles an evidence assessment mechanism can play is helping to establish greater consistency in how efficacy is understood and communicated. This includes creating clearer expectations around different forms of evidence, supporting stronger alignment between claims and evidence, and developing a shared language that enables more meaningful conversations about efficacy across the sector.

In doing so, it helps move the sector away from debates about the quantity of evidence and towards more sophisticated discussions about credibility, relevance, limitations and appropriate interpretation. More broadly, the pilot demonstrates the value of establishing a shared language for discussing evidence across the education sector. Through its criteria, guidance, reviewer training, educator resources and public outputs; the EEB has begun to define more consistent expectations around the evidence quality and efficacy. This shared understanding may prove to be one of the EEB's most significant long-term contributions to the system.

Recommendation 3: Consider further adaptations to the evidence assessment criteria and review process to reflect the rapid growth of AI-enabled products within the marketplace.

The rapid emergence of AI-enabled EdTech introduced a distinct challenge within the EEB pilot. The pace of product development and deployment in this area has, in most cases, outstripped the generation of robust evidence relating to impact. As a result, it was not feasible to expect products with significant AI components to demonstrate established efficacy within the same parameters as more mature products. Consequently, the initial

review pathway was selected as an appropriate mechanism through which to engage these products in the EEB process, allowing for early-stage examination of how AI-enabled products align with some elements of the core criteria, while recognising that evidence of impact (particularly in relation to criterion 6) would likely be under-represented at this stage.

Through this process, a number of areas for future consideration were identified. In particular, there is a need to explore how existing evidence assessment criteria developed during the pilot may require adaptation to account for the specific characteristics of generative AI systems. In particular, how evidence relating to the quality assurance of AI-generated content is captured and assessed, alongside broader consideration around transparency, consistency and reliability. As the marketplace continues to develop rapidly there is a need to consider potential mechanisms through which a future iteration of the framework could support and accelerate the generation of high-quality evidence in this space. . When evidence is evolving quickly, it may be less helpful to evaluate portfolios of evidence which have been gathered over time as much of this evidence may be out of date or relate to an earlier iteration of a product. It may be more helpful to look at how an evidence assessment mechanism might create efficiencies in the review process to enable it to review individual items of evidence and publish outcomes at a quicker pace. In this way, EdTech providers could submit evidence as it is generated, and evidence is timestamped to indicate exactly when, and which version of a product, the evidence relates to. An important focus for future development of this work should be to carry out exploratory work to examine how best to evidence efficacy of generative AI products, including how to create transparency around which large language models are being utilised and how they are trained for pedagogical use.

There may be opportunities to explore this further by coordinating with wider DfE initiatives such as the EdTech Testbed Programme and AI Tutoring Tools Pioneer Programme.

Conclusion

This report brings together insights drawn from two phases of the EdTech Evidence Board project incorporating the practical delivery of the EdTech Evidence Board (EEB) pilot, system-wide guidance and the contributions of a wide range of stakeholders across the education and EdTech ecosystem.

From these insights, phase two progressed the Edtech Evidence Board from a concept to a tested model. The pilot has demonstrated that an evidence-led approach to evaluating EdTech is both feasible and valuable, but that its effectiveness depends on recognising the current realities of the system in which it operates. While there is clear appetite across the EdTech sector to engage with evidence, the pilot has shown that capability, capacity and evidence literacy remain uneven. As such, a future evidence assurance system should perhaps function not only as a mechanism for assessment, but as a driver of greater coherence, clarity and maturity in evidentiary practice.

The pilot has provided strong validation of the core design principles established in phase one. The core criteria and review process are workable and capable of surfacing meaningful insights about product efficacy. Crucially, the process itself has proven to be a valuable intervention: enabling participating organisations to more clearly articulate their claims, understand the strengths and limitations of their evidence and identify tangible next steps for development.

The pilot provides strong evidence that relying on assessment alone to raise evidentiary standards is a limited intervention given the current state of evidence practice across the EdTech sector. The barriers identified - ranging from limited methodological expertise and the influence of market incentives - point to the need for a broader, system-level approach. This includes strengthening both the supply of high-quality evidence and the demand for it within education settings.

Phase two has helped to answer critical questions about how an evidence assessment mechanism might operate in practice, but it has also surfaced new and important lines of enquiry that must be addressed to enable further scaling, particularly in relation to evolving to better meet the needs of the education sector, whilst also meeting EdTech providers where they are, and bringing them on the journey.

A future phase of this project presents an opportunity to move from a successful pilot to an evolved, scalable sustainable infrastructure that supports the generation, interpretation and use of high-quality evidence across the system.

Acknowledgements

The project team remains hugely grateful for the time, openness and expertise offered by all those who have engaged with this phase of work, and whose input has been instrumental in shaping both the process and the learning captured in this report. We are particularly grateful to the EdTech organisations who were active participants in the pilot and have ultimately demonstrated their alignment to developing high standards of evidence in EdTech and whose practice has helped to shape them; to the members of the appointed board, whose oversight, scrutiny and considered judgement have played a critical role in ensuring the integrity and credibility of the review process, as well as providing crucial insight into the themes that inform our future recommendations; and to the members of the advisory and steering groups, whose expertise across education, research, accessibility and the EdTech sector has helped to shape both the design and refinement of the model.

We would also like to acknowledge the reviewers and panel members, whose expertise and rigour underpinned the assessment of evidence portfolios and the generation of high-quality outcome reports; and those who have contributed to the development of the EEB resources and capacity-building work, including the EdTech logic model tool and researcher-matching programme. These contributions have been vital in supporting organisations to engage meaningfully with evidence and in strengthening capability across the system.

Finally, we extend our thanks to the wider education community, including teachers, school and college leaders and practitioners who have taken the time to share their perspectives and experiences throughout the project.

We would like to formally acknowledge the following individuals and organisations for their contribution to this work:

The appointed board

- Dr Fiona Aubrey-Smith, Independent consultant and researcher
- Dr Robin Bevan, Education leadership specialist
- Kristy Evers, Education researcher and evaluator, Director at ImpactEd Evaluation
- Laurie Forcier, Senior Education Expert and VP Strategy at EDT&Partners
- Neelam Parmar, Education consultant and advisor, Director of Professional Learning at AISL Harrow
- Dr Kristen Weatherby, Education founder and leader, Founder of Breakthrough Labs
- Caroline Wright, Director General at BESA

Contributing individuals and organisations

- Adaptify Education
- Professor Alison Clark-Wilson
- All Aboard Learning
- Anneliese Smith
- Antoinette Hamilton
- Antonia Clary
- AQA
- BESA
- Beth Havinga
- Bow House Digital
- Caroline Wright
- Dr Cat Scutt
- Cheryl Shirley
- Christian Turton
- Daisy Education
- Daniel Davies
- Dave Marsh
- David Monis-Weston
- EdTech Impact
- Educake
- Education Quizzes
- Emily Yeomans
- Everway
- Dr Fiona Aubrey-Smith
- Dr Fiona Scott
- Gemma Gwilliam
- Genio
- Good Future Foundation
- Google for Education UK
- Hajar Bellazrak
- Dr Hayley Yates
- Heather Clements-Wheeler
- Helen Barker
- Hermione Thompson
- Holly Gardner
- James Reid
- Jérôme Nogues
- Jonty Lesse
- Joseph Arday
- Joysy John
- Kami
- Kathryn Taylor
- Dr Kristen Weatherby
- Kristy Evers
- Dr Laura White
- Laurie Forcier
- Learning by Questions (LbQ)
- Libby Hills
- Dr Lisa-Maria Müller
- Dr Lucy Rycroft-Smith
- Dr Lynne Beveridge
- Micheal Forshaw
- Micheal Taylor
- Neelam Palmer
- Oak National Academy
- Pobble
- Renate Samson
- Dr Robin Bevan
- Rohan Slaughter
- Sapio
- Scott Hayden
- Sean Harris
- Showbie
- Simon Hay
- StepsWeb
- techUK
- Teacher Success Platform
- Teachmate
- The Quotation Bank
- Running Paper
- Professor Tom Crick
- Valerie Handunge
- Victoria Hedlund

Appendix A: EdTech Evidence Board evaluation criteria (including subcriteria)

Criterion 1: Research/evidence informs product evaluation and ongoing development

- 1 Evidence gathering is planned for as part of ongoing evaluation and development activity (1.1)
- 2 Evidence is gathered from multiple sources [for example, this may include research literature, findings from experimental and non-experimental research, and analysis of data sources] (1.2)
- 3 Evidence is gathered from, and in collaboration with, education settings [e.g. to evaluate product effectiveness, understand impact and inform ongoing development] (1.3)
- 4 Learning from research and evidence is incorporated into iterative product development (1.4)
- 5 Evidence is gathered to evaluate the impact of improvements once implemented (1.5)

Criterion 2: Research/evidence indicates there is a sound educational need for this product

- 1 A logic model or theory of change specifies intended impact(s) on teachers and/or learners (2.1)
- 2 The identified educational need for the product is aligned to published curriculum or wider educational goals (2.2)
- 3 Published theory, literature or research indicates the product has the potential to fulfil a plausible educational need (2.3)
- 4 Market research, user research or user testing indicates the product has the potential to fulfil a plausible educational need (2.4)
- 5 Research gathered about the product itself indicates that the product fulfils a plausible educational need (2.5)

Criterion 3: The design of the product is informed by evidence

- 1 A logic model or theory of change demonstrates how the product seeks to enhance teaching and/or learning to achieve its intended impacts (3.1)
- 2 The pedagogical design of the product, including content and learner activities where relevant, aligns with theory and evidence from education and the wider learning sciences (3.2)
- 3 The pedagogical design, including content and learner activities where relevant, utilises specific choices that are grounded in robust research and evidence [e.g. underpinning ideas have a well-agreed theoretical basis, are drawn from substantiated research or synthesised from findings that have been replicated across multiple studies] (3.3)
- 4 The product has been designed to support teachers to develop digital competencies or digital pedagogy skills (3.4a) or the product has been designed to

- support learners to develop digital literacy or digital citizenship skills (3.4b)
- 5 The design of the product is informed by learning from research and evidence gathered about the product itself (3.5)

Criterion 4: Data from users indicates that the product is able to be utilised effectively in schools and/or colleges

- 1 Analytics are used, where available, to help understand user experiences (4.1)
- 2 Analytics data and/or feedback from users helps to build a picture of the strengths and limitations of the product (4.2)
- 3 Feedback from users indicates that the product supports pedagogical needs effectively (4.3)
- 4 Evidence about user experiences is gathered from users in a range of contexts (4.4)
- 5 Data and/or user feedback have been used to improve user experience (4.5)

Criterion 5: Research/evidence demonstrates that the product can enable or enhance teaching and/or learning

- 1 Theory or wider research suggests the product may have potential to enhance teaching and/or learning in one or more specified ways (5.1)
- 2 Evidence gathered shows that the product enables or enhances teaching and/or learning in one or more specified ways (5.2)
- 3 Evidence gathered demonstrates how the product can be incorporated effectively into classroom practice (5.3)
- 4 Research/evidence helps to build an understanding of the factors that may contribute to, or impede, effective implementation (5.4)
- 5 Evidence gathered reflects a range of contexts and users [including where relevant, learners with Special Educational Needs or Disabilities or those recognised as being educationally disadvantaged/underserved] (5.5)

Criterion 6: Evidence for the product demonstrates positive effects on identified outcomes for learners and/or teachers

- 1 Evidence demonstrates positive impacts on identified outcomes for teachers [for example, on teacher workload, teacher self-efficacy, teaching skill, digital competency] (6.1a) or Evidence demonstrates positive impacts on identified outcomes for learners or specified groups of learners [for example, on student attainment, cognition and learning, skills development, motivation and engagement, digital literacy, confidence or self-efficacy, mental health/wellbeing] (6.1b)
- 2 Evidence has been sought to understand the extent to which teachers or learners may experience potential negative impacts [for example, on teacher workload; or on learner cognition and learning, motivation and engagement, mental health/wellbeing] (6.2)
- 3 Evidence gathered over time demonstrates evidence that the product can achieve its intended long-term impacts (6.3)
- 4 Evidence demonstrates positive impacts across a range of contexts (6.4)
- 5 Evidence demonstrates positive impacts for pupils who face educational

disadvantage, helping to narrow attainment gaps (6.5)

© Crown copyright 2026

This publication is licensed under the terms of the Open Government Licence v3.0, except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

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

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