Understanding the quality of EdTech interventions and implementation for disadvantaged pupils

Protocol for a systematic review with meta-analysis

Open Development & Education

May 2024

https://doi.org/10.53832/opendeved.1077

Education Endowment Foundation

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus – Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report, please contact:

Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
0207 802 1653
info@eefoundation.org.uk
www.educationendowmentfoundation.org.uk

Recommended citation: Haßler, B., McBurnie, C., Walker, H., Klune, C., Huntington, B., & Bhutoria, A. (2024). *Understanding quality characteristics of EdTech interventions and implementation for disadvantaged pupils: Protocol for a systematic review with meta-analysis*. Open Development & Education. https://doi.org/10.53832/opendeved.1077. Available under Creative Commons Attribution 4.0.

Contents

Figures and tables

1. Background and rationale for the review

In 2019, the Education Endowment Fund Foundation (EEF) commissioned an evidence review on the impact of educational technology — 'EdTech' — on student attainment. The review highlighted the challenge of keeping abreast of research developments in the field of EdTech, which continues to evolve rapidly (↑Lewin et al., 2019). Since the publication of the review, teachers across the world have been required to turn to technology to ensure learning continuity for over 1.6 billion out-of-school learners during lockdown (↑Karboul, 2020). In England specifically, over two-thirds of schools have introduced, upgraded, or increased their use of technology in the 2020s, especially since the COVID-19 pandemic (↑CooperGibson Research, 2022).

This was preceded in the 2010s by significant growth in the EdTech sector, with schools and teachers increasingly making use of EdTech in classrooms. The COVID-19 pandemic further catalysed this growing use of EdTech. In the aftermath of pandemic-induced school closures, teachers are now more often expected to select and implement new EdTech models, but often without guidance on what works for what students (↑Pokhrel & Chhetri, 2021; ↑West et al., 2023). At present, the evidence base on educational technology presents mixed messages about when and how to use technology in the classroom. As such, especially in the wake of the pandemic and widespread implementation of different EdTech, there is an urgent need to better understand how EdTech functions in the classroom, and in particular for those learners from disadvantaged backgrounds. For example, existing research indicates that disadvantaged students can benefit more from technology than advantaged students (↑McNally et al., 2016; ↑Takacs et al., 2015). At the same time, evidence indicates that the use of EdTech can exacerbate learning inequality as disadvantaged students have less access to hardware and software (↑Vicentini et al., 2022).

To date, limited reviews have systematically examined why and how EdTech interventions succeed or fail in different contexts. For example, ↑Outhwaite et al. (2023) showed that mobile applications offering a personalised learning journey with explanatory and motivational feedback had the greatest impact on numeracy outcomes for children aged 0–8 years. Meanwhile, ↑Verbruggen et al. (2021) and ↑Outhwaite et al. (2023) highlighted a broader range of implementation features — parental support, frequency of intervention, and professional development — associated with effective EdTech interventions. However, these reviews focused on either a specific technology (e.g., mobile applications) or a specific subject domain (e.g., mathematics).

In this context, we are working with EEF to analyse how the use of EdTech can raise student attainment across all subject domains such as literacy, numeracy, English (as a modern foreign language), science, and ICT / computing, with a focus on disadvantaged pupils. Our preparatory work — the initial stakeholder engagement outlined in Phase 1 below — demonstrates that the impact of technology on student attainment is not determined by the use of a specific device or software. Instead, it is how these devices and software are used to promote learning, and what supportive factors are in place, that matters (↑Clark et al., 2015). This finding underscores the need for a detailed analysis of the mechanisms that underpin effective EdTech interventions.

To conduct this analysis, we will employ an innovative and sequential mixed-methods research design that synthesises different research methods in three distinct phases:

- **Phase 1.** Through initial stakeholder engagement, education researchers, school leaders and teachers will be consulted to define the conceptual framework for this project. In doing so, we aim to ground our study in the lived experience of education practitioners in England (est. March–May 2024).
- **Phase 2.** We will systematically review current literature on the implementation and impact of EdTech interventions, focusing particularly on the impact on disadvantaged students. We will screen studies for relevance and quality, extracting key evidence from high-quality studies and coding for the 'mechanisms' that shape programme impact and conduct a meta-analysis (est. May–October 2024).
- **Phase 3.** Our findings will be presented to education practitioners, and we will invite evidence-based insights and practice-based critiques as part of a structured community review. Through this innovative process, we will refine and validate our proposed mechanisms and their relationship to the learning outcomes of disadvantaged pupils (est. October–November 2024).

This approach will enable us to manage complex data while maintaining alignment with the realities of practitioners. The addition of a structured community review (Phase 3) will allow us to address the methodological limitations of systematic reviews. The overall methodology and specifics of each Phase are detailed in Section 4 on methodology.

2. Objectives

For this project, we will conduct a systematic review with meta-analysis. Three primary objectives will be addressed to understand the reasons why EdTech interventions succeed or fail:

1. Identify the core mechanisms of EdTech interventions that lead to improved attainment outcomes for pupils.
2. Explore intermediate outcomes associated with increased pupil attainment for EdTech interventions and their placement within mechanisms of change.
3. Investigate potential differences in impact related to various mechanisms within EdTech interventions, with a focus on disadvantaged pupils.

As an organisation, Open Development & Education is committed to producing open-source, accessible research that furthers shared understanding and good practice. As a result, we have two secondary objectives:

4. Produce an easily accessible, interactive evidence map that builds on facilities within the EPPI Reviewer ecosystem (e.g., ⇡Bond, 2020; ⇡Bond et al., 2020);
5. Document and publish guidance on the research process implemented for replication in future projects.

To achieve these objectives, we will examine a primary research question (RQ):

> **What mechanisms of EdTech interventions are associated with improved pupil attainment outcomes?**

In doing so, we will explore the following sub-research questions.

1. What mechanisms are identified in school-based EdTech intervention studies?
2. What is the impact of school-based EdTech interventions on attainment outcomes in countries with ‘high technological readiness’?
3. What intermediate outcomes are associated with improved pupil attainment in EdTech interventions, and where do they fit within the identified mechanisms?
4. What is the differential impact of EdTech interventions on pupil attainment based on socioeconomic status?
5. Are there differences in impact associated with different mechanisms?
6. What are the key characteristics of school-based EdTech intervention studies implemented in countries with ‘high technological readiness’?

The final research report will be published on the EEF website, and potentially used to inform future EEF grant-making and teacher guidance.

3. Definition of key terms

This section defines EdTech, mechanisms of change, and provides our definition of 'disadvantage'.

3.1. Defining educational technology

Educational technology (EdTech) is an umbrella term encompassing a range of technologies and meanings interpreted differently by different people (↑Outhwaite et al., 2023). In this light, previous reviews have employed different definitions and scopes to examine the impact of EdTech on student attainment.

A common approach has been to focus on a specific device or approach. For example, several systematic reviews have focused on the role of interactive multimedia in improving reading skills and STEM outcomes (↑Abrami et al., 2020; ↑D'Angelo et al., 2014; ↑Takacs et al., 2015). Other reviews have explored the effectiveness of different approaches to online, blended, and computer-assisted learning (↑Kunkel, 2015; ↑Lin, 2014; ↑Means et al., 2013; ↑Sokolowski et al., 2015; ↑Sung et al., 2016; ↑Zheng et al., 2016; ↑Zheng et al., 2018). Several studies have also analysed personalised and adaptive learning programmes, including intelligent tutoring systems (↑Belland & Belland, 2017; ↑Belland et al., 2017; ↑Gerard et al., 2015; ↑Kulik & Fletcher, 2016; ↑Major et al., 2021; ↑Steenbergen-Hu & Cooper, 2013; ↑Zheng, 2016). A notable subset of this work has reviewed the evidence relating to educational applications and game-based learning (↑Chen et al., 2018; ↑Clark et al., 2016; ↑Outhwaite et al., 2023; ↑Outhwaite et al., 2023; ↑Wouters et al., 2013). Established forms of artificial intelligence (AI) are a well-researched area of EdTech; however, with the recent development and application of generative AI tools, most notably LLMs/ChatGPT, this body of research has only recently begun to examine generative AI applications and impacts in educational contexts (↑BaiDoo-Anu & Owusu Ansah, 2023; ↑Holmes & Tuomi, 2022; ↑Zhai et al., 2021).

However, evidence shows that choices on how to use technology in the classroom have a greater impact on learning than the choice of device or approach (↑Higgins et al., 2012; ↑Lewin et al., 2019; ↑Verbruggen et al., 2021; ↑Walker et al., 2023). ↑Archer et al. (2014) provided the opportunity to further explore this by examining the impact of technology in three roles: a tutor, a teaching aid whereby media (e.g., interactive activities and videos) is used to enhance teacher-led instruction, and a learning tool where students use technology independently to enhance their learning. More recently, ↑Lewin et al. (2019) evaluated the use of technology to support learning from experts, with others, through making, through exploring, through inquiry, through practising, from assessment, and in and across settings.

This study will adopt a deliberately broad definition of EdTech:

> *"Education technology (EdTech) refers to the practice of using technology to support teaching and the effective day-to-day management of education institutions. It includes hardware (such as tablets, laptops, or other digital devices), and digital resources, software and services that help aid teaching, meet specific needs, and help the daily running of education institutions (such as management information systems, information sharing platforms and communication tools)"* (↑UK Government, 2019: p. 5).

Under this definition, EdTech could refer to any device, hardware, or digital approach that supports teaching and learning activities such as lesson delivery, group work, and assessment (↑CooperGibson Research, 2022).

For the purpose of this review, we will focus on interventions and approaches that target students and aim to improve student attainment, using quantitative measures in any curriculum subject. In doing so, we will exclude interventions and approaches that support teacher professional development and school administration, and those that solely focus on the provision of hardware.

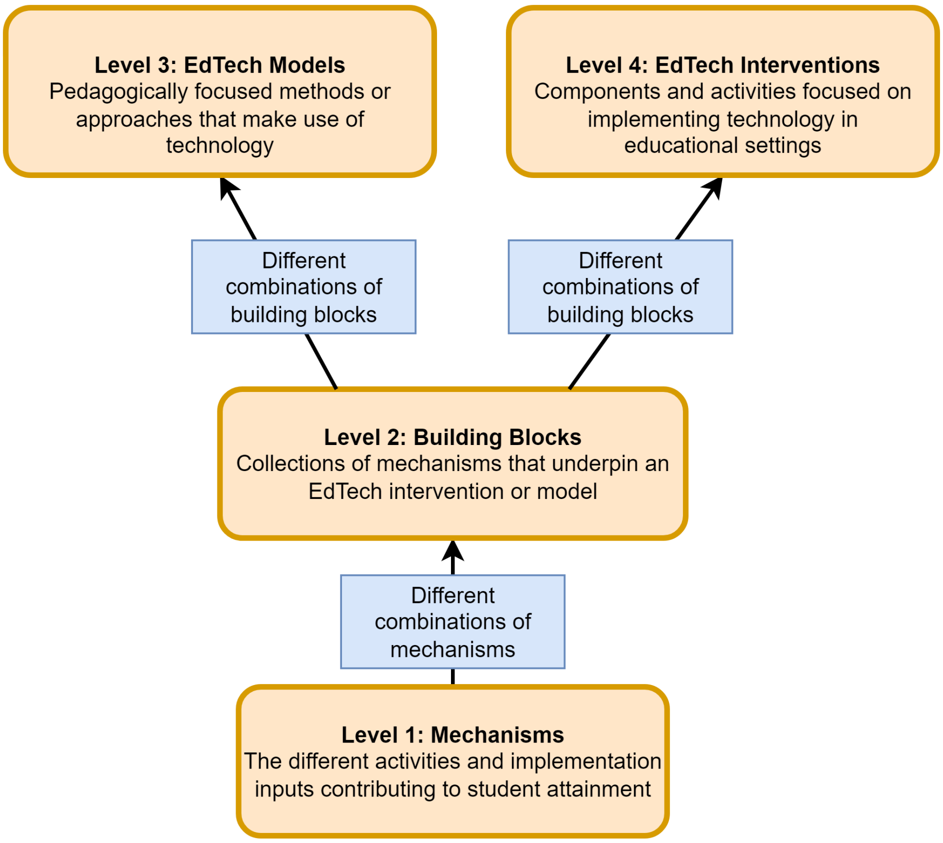
3.2. Defining our conceptual framework

Building on prior research, this review aims to identify the mechanisms that underpin effective EdTech interventions. Identifying mechanisms addresses current gaps in the limited systematic reviews of EdTech interventions in providing a needed qualitative element to support and deepen explanations of quantitative studies. When evaluating different interventions and models, researchers often consider randomised controlled trials (RCTs) as the gold standard (↑Hariton & Locascio, 2018). However, impact evaluations are unable to distinguish causally important features of evidence-based interventions from causally redundant features (↑Sims et al., 2021). As a result, these studies provide decision-makers with little information on how to replicate an intervention in another context and teachers with little guidance on how to effectively use technology to support learning in different classrooms (↑Alison, 2023; ↑Moore et al., 2015; ↑Williams, 2020). Furthermore, the heterogeneity of impact estimates in education evaluations compound this challenge. For example, studies have reported significant variation in the impact of similar interventions that have been delivered in the same context at the same time (↑Bold et al., 2018; ↑Evans & Popova, 2016; ↑Kerwin & Thornton, 2021). These interventions achieved inconsistent results, as small modifications to the underlying model influenced if and how different inputs worked alongside each other (↑Kerwin & Thornton, 2021). This conclusion aligns with international research on complementarities between inputs in education programmes (↑Glewwe et al., 2016; ↑Mbiti et al., 2019; ↑Piper et al., 2015). In light of these problems, the identification of mechanisms can serve to better contextualise and enhance explanations among quantitative studies of EdTech interventions, especially as it relates to student attainment.

For this purpose, we will examine EdTech interventions at four levels: mechanisms, building blocks, interventions, and models (Figure 1). These different levels are to be viewed as interconnected, to form a complex ecosystem where mechanisms, building blocks, interventions, and models dynamically interact and potentially impact student learning outcomes. While our research questions focus solely on identifying mechanisms as part of this study, this conceptual framework acknowledges that multiple mechanisms can be present within different levels and arrangements of EdTech interventions. While this study may not identify specific building blocks, models, and interventions, they are still conceptually needed in order to properly examine and identify the specific mechanisms within them. This is especially important for the Phase 2 systematic review, as literature collected and analysed may focus or present EdTech in different ways, which our conceptual framework attempts to capture.

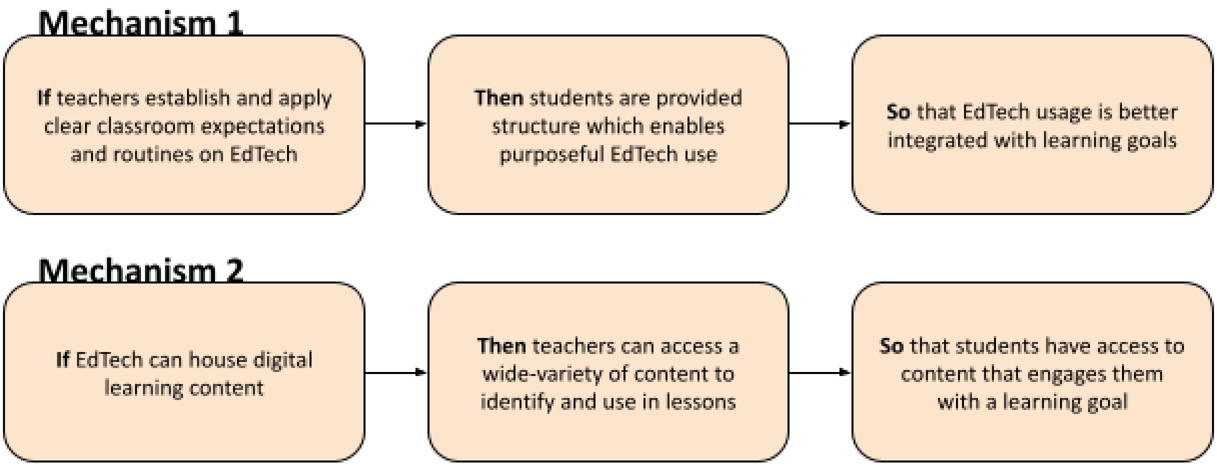
Definitions for each level of our conceptual framework are presented below, with examples to help illustrate the concepts. However, not all of the elements and examples mentioned within the definitions are within the scope of the current review. For detailed information on what is included within this review, please refer to the inclusion criteria.

Figure 1: Conceptualisation of EdTech mechanisms, building blocks, models, and interventions



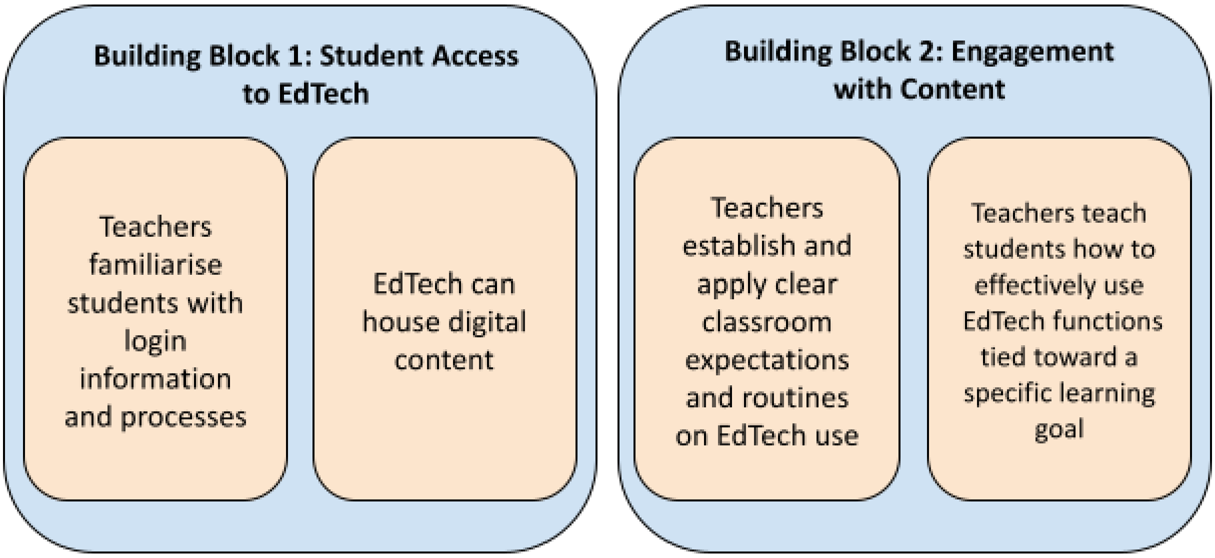
We will focus on identifying the **mechanisms (Level 1)** that underpin impactful use of EdTech. In doing so, we will adopt and extend the definition of 'mechanisms' by ↑Illari & Williamson (2012) as the *"entities and activities organised in such a way that they are responsible for the* *phenomenon".* This definition was similarly adopted in a previous EEF-supported study on the characteristics of effective teacher professional development (↑Sims et al., 2021). The "phenomenon" that our study is concerned with is improved student attainment. As such, the "entities and activities" of mechanisms refer to the inputs that characterise the use and implementation of EdTech. The entities could include the software or hardware utilised as part of an EdTech intervention, as they are tangible tools that contribute to the implementation. Activities would include any practices, behaviours, or other activities that are employed that contribute to improved student attainment, such as those inputs indicated by the 'If' statements in the example mechanisms in Figure 2. Importantly, this approach builds on definitions of 'mechanisms' in the fields of education, economics, public health, and public management (↑Barzelay, 2007;↑Bates & Glennerster, 2017; ↑Harn et al., 2013; ↑Ludwig et al., 2011; ↑Leviton, 2017; ↑Moore et al., 2015; ↑Sharples et al., 2019; ↑Sims et al., 2021). Building on this previous work, this study views that a mechanism is generally a micro-level input that makes up the building blocks that define EdTech models and interventions. As such, mechanisms are likely to be seen at the classroom and school level (see Figure 2 for example mechanisms). Given this, our initial conceptualisation aims to provide much more descriptive detail, highlighting how the specific practice, behaviour, or activity work toward an outcome related to improved student attainment.

Figure 2: Example Level 1 mechanisms, pictured with an explanatory 'If… then… so' structure



Building blocks (Level 2) are a well-articulated collection of mechanisms that underpin EdTech interventions and models. In this sense, a building block might be considered as a specific aspect of an EdTech model or intervention that multiple mechanisms contribute to, or are required for. This framing recognises that specific building blocks might be made up of different mechanisms that can work together in different ways and at different times to create different learning outcomes (↑Barzelay, 2007; ↑CooperGibson Research, 2022; ↑Outhwaite et al., 2023). Using a flipped classroom model as an example, Figure 3 shows building blocks (denoted by the blue shape) that define the specific aspects of the flipped classroom model that is made up of different mechanisms (denoted by the orange shape).

Figure 3: Level 2: building blocks with collected mechanisms



EdTech models (Level 3) describe pedagogically driven methods or approaches that make use of technology. In this sense, teachers or other educational practitioners may implement different EdTech models in their teaching practice with the application of technologies. An example of an EdTech model is a flipped classroom where students engage with instructional content online before class sessions (↑Akçayır & Akçayır, 2018; ↑Cabı, 2018). Similarly to interventions (described below), EdTech models are made up of different building blocks (level 2) and mechanisms (level 1) that make up the scope, purpose, and activities that contribute to improved student attainment.

EdTech interventions (Level 4), sometimes also referred to as programmes, differ from EdTech models in that an intervention centres around the set of activities that focus on the implementation of a technology in an educational setting, which can include pedagogical (EdTech directly related to teaching, such as an app that provides maths games) or non-pedagogical (EdTech related to non-teaching purposes, such as a learning management system) interventions. As such, interventions consist of hardware, software, other digital approaches, or a combination of these components alongside a defined set of activities and materials aimed at implementing technology in educational settings to achieve specific goals or outcomes. In this conceptualisation, interventions prioritise technology use for a variety of educational purposes. Examples of EdTech interventions may include the implementation of a learning management system (LMS) in a school, the use of educational apps for language learning, or the use of adaptive learning software to personalise mathematics instruction. A real-world example of an EdTech intervention is the non-profit organisation *onebillion's* 'onecourse' programme that uses adaptive software to deliver personalised reading, writing, and numeracy lessons to primary school children (↑Outhwaite et al., 2017; ↑Outhwaite et al., 2019; ↑Outhwaite et al., 2023). As with EdTech models (Level 3), EdTech interventions are made up of different building blocks (Level 2) that underpin the scope and purpose of the intervention. The building blocks themselves consist of the different mechanisms (Level 1), which are the possible activities and implementation inputs of the intervention that contribute to improving student attainment.

As part of our analysis, we will examine if and how contextual variables mediate the impact of mechanisms in different settings (↑Barzelay, 2007; ↑Castro et al., 2010; ↑Harn et al., 2013; ↑Leviton, 2017; ↑Moore et al., 2015). In doing so, we recognise that the same building blocks and mechanisms lead to different behavioural responses in different places and, therefore, achieve varying levels of impact across contexts (↑Bates & Glennerster, 2017; ↑Castro et al., 2010; ↑Pritchett & Sandefur, 2015; ↑Verbruggen et al., 2021;↑Vivalt, 2020; ↑Williams, 2020).

This conceptual framework reflects the research team's current framing of mechanisms within the use of Edtech. We acknowledge that as the research progresses, new insights may emerge that lead to the adjustment of the framework and the relationship between different components. This framework should be considered an iterative tool that evolves alongside the research process, and will be reviewed and updated accordingly.

3.3. Defining disadvantage

While this review will search broadly for school-based EdTech interventions with children, the main focus of interest is on EdTech interventions for disadvantaged children.

For the purpose of this review, we will use the ↑UK Government's (2023) definition of disadvantaged children. Based on this definition, disadvantaged children in the UK are:

- recorded as eligible for free school meals, or have been recorded as eligible for free school meals in the past 6 years;
- looked after by a local authority, or previously looked after by a local authority.

In 2023, the proportion of disadvantaged pupils ranges from 26% of students in Key Stage 1 and Key Stage 4 to 30% of students in Key Stage 2 (↑ONS, 2023a; ↑ONS, 2023c). Importantly, only 25% of disadvantaged pupils in Key Stage 4 achieve Grade 5 and above in English and mathematics compared to 45% of non-disadvantaged pupils (↑ONS, 2023c). A similar trend exists across all other levels of the UK education system (↑ONS, 2023a; ↑ONS, 2023b; ↑ONS, 2023d).

At the same time, we will rigorously code included studies to highlight factors that intersect with economic disadvantage and are associated with low attainment among students in Key Stages 1–5. In doing so, we recognise the multidimensionality of disadvantage and acknowledge that the above factors intersect to compound other forms of marginalisation.

For example, children with special educational needs have learning difficulties or disabilities that make it harder for them to learn than most children of the same age. In 2023, the proportion of children with a reported special educational need ranged from 17% of students at the end of Key Stage 4 to 20% of students at the end of Key Stage 2 (↑ONS, 2023b;↑ONS, 2023c). At all levels, pupils with special educational needs have significantly lower attainment than pupils without special educational needs. At the end of Key Stage 2, for example, 20% of pupils with special educational needs achieve the expected standard in reading, writing, and maths compared to 66% of other pupils (↑ONS, 2023c).

Furthermore, student attainment varies significantly across ethnic groups. Notably, Roma pupils, Traveller of Irish Heritage pupils, and Black Caribbean pupils consistently ranked as the lowest-performing ethnic groups in Key Stages 1–4 (↑ONS, 2023a; ↑ONS, 2023b; ↑ONS, 2023c). In Key Stage 1, for instance, only 23% of Roma pupils met the expected standard in reading (↑ONS, 2023a).

Geographically, relatively small regional differences in student outcomes masked large variations in school performance across local authorities. In Key Stage 4, the proportion of pupils achieving 5 or above in English and mathematics ranged from 19.3% to 69.1% across different local authorities (↑ONS, 2023c). Across all levels, pupil attainment was typically lowest in local authorities in northern, midland, and coastal regions.

In summary, for the purpose of this study and sections of the analysis concerning economically disadvantaged pupils, we will code for and focus on those that fall within either of the following categories:

- eligible for free school meals, or have been recorded as eligible for free school meals in the past 6 years;
- under the care of a local authority, or previously under the care of a local authority.

Meanwhile, while our analysis will not focus on these categories, we will also code studies to highlight the following additional forms of disadvantage:

- exhibits special educational needs or difficulties that make it harder to learn than most children of the same age;
- identifies as Roma, Traveller of Irish Heritage, or Black Caribbean;
- attends school in a low-performing local authority.

Coding for these categories will be beneficial when combining the studies with the existing EEF database, improving the ability to sort, filter, and conduct future analyses.

4. Methodology

Overall, we will use a sequential mixed-methods approach to examine whether there are mechanisms of EdTech interventions that improve attainment outcomes for pupils — and, especially, disadvantaged pupils. The approach has three phases:

Phase 1. Stakeholder engagement

Phase 2. Systematic literature review with meta-analysis

Phase 3. Structured community review of outcomes

The Phase 1 stakeholder engagement with researchers and school practitioners enables a more accurate development and framing of the Phase 2 systematic literature review protocol, ensuring that the work is aligned with insights of practitioners. Next, we will conduct a systematic literature review with meta-analysis to provide a quantitative synthesis of the impact of different uses of EdTech on student attainment. This will allow us to identify empirical examples of effective uses of EdTech. We will then pinpoint the 'building blocks' and the 'mechanisms' of interventions that improve attainment outcomes.

Lastly, in Phase 3, a structured community review will be conducted. A structured community review is a process that includes sets of artefact-based semi-structured interviews, focus groups, and online surveys that serve as structured reviews of a draft report (the artefact), with a view to determining the external validity of the draft report, i.e., in our case the meta-analysis. This is an approach we have innovated in the context of education research (↑Haßler, 2021) that addresses the methodological issue of external validity, common to both meta-analysis and RCTs. In education research, these gaps are characterised by a focus on overall programme effectiveness (quantitative, excessive focus on internal validity) while providing little insight as to the nuances of why an intervention did or did not work when applied in different contexts (mixed methods, with consideration for external validity; also see ↑Pawson et al., 2005).

For this structured community review, we will once again consult a community of education practitioners (consisting of teachers, teaching assistants, specialist teachers, and educational psychologists) with the results of the Phase 2 systematic review. In this way, we will be able to test, validate, and further develop insights on effective EdTech use from Phase 2 against the lived realities of practitioners. We seek to assess how these insights may or may not reflect mechanisms that emerged from the systematic review, providing grounding and assessment of feasibility. Through the community review, we also aim to further assess how the findings from the systematic review apply to disadvantaged pupils, by identifying mechanisms that might be particularly important for this group.

4.1. Phase 1: Stakeholder engagement

As an organisation, we are committed to grounding our work in the expertise and lived experiences of education practitioners in England. As such, we will collaborate with researchers and teachers to define the scope and focus of the project in two stages before reviewing any literature.

4.1.1. Stage 1. Consulting education researchers

In the first stage, we engaged in a critical friend review with three education researchers, prominent either in the EdTech research field and / or the methodology that we will be using in this research. The choice of researchers was strategic and selected in collaboration with the EEF to ensure that we get relevant, expert feedback on our chosen approach. This critical friend review was designed to reflect that of peer review in research. During these discussions, researchers were presented with the definitions of key terms and conceptual framework, as the primary aim is to seek validation and refinement of our initial conceptualisation through their expert feedback. In these interviews, we:

- explained the purpose and scope of the project;
- presented our methodological approach;
- used our draft definition of key terms as an artefact to discuss our definitions of EdTech, conceptualisation of mechanisms of change, and definition of disadvantage;
- solicited feedback to refine our methodological approach, definitions, and conceptual framework.

To facilitate these discussions, participants received a written overview of our methodological approach, definitions of key terms, and conceptual framework in advance. Key learnings and iterations gained from these collaborative discussions will be included in the final report as an annex. These discussions were conducted on a digital platform such as Zoom or Google Meet. Any feedback received from this consultation process was discussed with the research team and necessary changes were made to incorporate the feedback to improve our methodological approach and definitions before presenting these to the teachers and support staff in Stage 2. This process took place in February, following our initial conceptualisation phase.

4.1.2. Stage 2. Teacher interviews and surveys

In the second stage, we organised 45-minute semi-structured interviews and online surveys with teachers working in schools in England. The main aim is to strengthen the relationship between our conceptual framework and real classroom practice, in addition to broadening the keyword inventory for the systematic review. In these interviews, we will:

- explain the purpose and scope of the project;
- present our definitions of EdTech, mechanisms of change, and disadvantage;

- invite participants to critique, edit, and expand the definitions (as needed);
- discuss the core categories and sub-categories for our keyword inventory;
- ask participants to list keywords under each category and sub-category.

The research team aims to conduct a minimum of 5 teacher interviews, with a mixed sample of primary and secondary level teachers, working at schools in different locations across the UK, and inclusive of senior level teachers. An interview schedule was developed by the research team and the final interview schedule can be found in Annex 1.

The teacher survey is primarily to capture teachers' uses and experiences of EdTech within the classroom. We will explain the scope and purpose of the project, and then ask questions to capture their teaching role, background with educational technology, and perspectives on the use of technology in the teaching and learning processes. The full survey can be found in Annex 2. The primary aim of the survey is to further inform our conceptualisation and methodology, but also provide initial ideas about how EdTech is used within a classroom setting, and perspectives on how this impacts learning. As with the interviews, the teacher survey was also refined in consultation with the EEF, before being widely distributed.

This survey will be open to all teachers in the UK and shared via social media channels of the EEF and Open Development & Education to ensure maximum engagement. The aim is to qualitatively capture a wide range of experiences and perspectives over the 4-week data collection period. As the responses will not be quantitatively analysed, but rather thematically analysed and used to map potential mechanisms during the conceptualisation phase of our research (Section 4.1.3), there will be no minimum threshold for responses. Each survey response will contribute meaningfully to the mapping process, as mechanisms will be extracted from the survey responses. This will allow us to gain a comprehensive understanding of different EdTech mechanisms used within the classroom.

The team will follow a careful approach, as laid out in the Data Management Plan, to adhere to important measures around ethics, consent, data storage, security, archiving, and so on.

4.1.3. Stage 3. Interpretation of results

Interviews and open-ended responses from surveys will be thematically analysed and inductively coded into broad themes that characterise teachers' perspectives and use of EdTech in schools. The coded thematic areas will then be further analysed to identify the specific practices, behaviours, and activities that make up mechanisms of effective EdTech use, barriers to effective EdTech implementation, technologies used by teachers, and teachers' perspectives on Edtech use with disadvantaged students. Mechanisms interpreted from the coded data will be recorded and mapped to produce a first iteration of a collection of potential mechanisms that have been identified by teachers. This initial mechanisms map will then be refined over the course of the project in Phases 2 and 3, as further described below. The additional areas of analysis, including the mechanisms, will be used to further inform and refine the project's keyword inventory in the Phase 2 screening of publications.

The overall findings will also be documented in an initial results report to be presented to the EEF. This report will summarise findings from the analysis in terms of respondent attitudes toward EdTech, key trends in EdTech use relating to factors connected to student learning, teaching practice, barriers to integrating EdTech, and respondents' views on disadvantaged students' interactions with EdTech.

There are clear limitations to basing our interpretations of mechanisms within the classroom on a small number of self-reported survey and interview responses, as they are not necessarily reflective of universal teacher experiences but rather concern a teacher's own personal experiences, which may differ significantly based on their individual school context and prior experiences with EdTech. However, as mentioned, we are committed to reflecting the lived experiences of practitioners. Furthermore, the identified limitations will be addressed by the following steps of interpreting within existing literature for Phase 1, then further building on teacher experiences with Phases 2 (existing literature) and 3 (a broader sample of experienced practitioners).

As such, these initial results will then be interpreted within the context of a collection of existing literature reviews and meta-analysis to refine the initial iteration of the mechanisms map outlining and categorising the specific practices, behaviours, and activities used in connection with EdTech to improving student learning outcomes.

The map generated from the data collected in Phase 1 surveys, interviews, and literature will be used to generate an initial list of mechanisms that is grounded in practitioner experiences and existing literature. Similar to the conceptual framework in Section 3.2, this mechanisms map will be iterative and will continue to be refined through the research process and results of Phases 2 and 3.

The results of the data extraction from Phase 2 will further refine the map. If there is no robust literature to corroborate a mechanism identified by teachers during Phase 1, the mechanism will be kept on the mechanisms map, as it was a lived experience shared in our research, but it will be noted that this is not currently supported by the reviewed literature.

This map will likely form one of the artefacts used for data collection in the Phase 3 community structured review. Ultimately, the results of the community structured review will then be used to identify and create a refined map of core mechanisms.

4.2. Phase 2: Systematic review and mixed-methods meta-analysis

Once the initial stakeholder engagement stage is complete, we will conduct a systematic review with meta-analysis to provide a mixed-methods synthesis of the impact of different EdTech interventions that improve attainment outcomes for students. In doing so, we will be able to identify empirical examples of effective uses of EdTech and pinpoint the building blocks and mechanisms of these interventions. The results of the systematic review and meta-analysis will then feed into the working mechanisms map from Phase 1, consolidating the existing mechanisms and adding mechanisms not previously identified that have been shown to drive improvement in attainment outcomes through EdTech interventions.

In addition to the traditional systematic review and meta-analysis process, we will produce an easily accessible interactive evidence map to provide a structured overview of existing research on EdTech interventions within this context (outlined further in Section 4.2.5). We will also consolidate, document, and publish the methodological process for enhanced transparency so that the process can be utilised and easily adapted by other researchers in the field. This makes the creation of a living review possible, which is compelling in this field given the rapidly evolving nature of evidence in EdTech. The dynamic nature of this field necessitates a flexible approach to evidence synthesis, allowing for timely updates and adjustments to reflect the latest research findings. Transparent documentation of our methodological processes will allow new teams of researchers to efficiently build on existing work rather than having to repeat basic steps.

Our methodology for conducting the systematic review and meta-analysis will adhere to established Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines to ensure methodological rigour throughout the review process (see Annex 3 for a template of the flow diagram that will guide our search and screening process). There will be 9 stages, with further descriptions in the corresponding sections:

1. Systematic searches
2. Validation of articles against search terms
3. Eligibility screening of titles, abstracts, keywords
4. Initial full-text manual screening
5. Evidence mapping
6. Full-text quality appraisal
7. Data extraction and management
8. Effect size calculation
9. Content analysis

Our approach to the meta-analysis is informed by the critique and recommendations of ↑Tipton et al. (2019a) and ↑Tipton et al. (2019b). ↑Tipton et al. (2019a) note several points of consensus, such as using all relevant effect sizes in a single model and dividing hypothesis tests into confirmatory and exploratory (pp. 173–174). We also agree with transparency requirements for making available underlying data in machine-readable formats (↑Tipton et al., 2019b). Overall, relevant methodological approaches and nuances will be considered (e.g., ↑Ahn et al., 2012; ↑McNeish, 2017; ↑Moeyaert et al., 2017; ↑Scammacca et al., 2014; ↑Sharpe & Poets, 2020; ↑Van den Noortgate et al., 2015).

4.2.1. Stage 1: Systematic searches

We will employ an automated, cross-database search strategy as 'best practice' for systematic literature reviews in the fields of education and EdTech. This is motivated by the observation that *"no database contains the complete set of published materials"* (⇡Xiao & Watson, 2017: p.11). Searching across multiple databases for literature is essential for conducting thorough and comprehensive research. By utilising multiple databases, we can access a wider range of sources, increasing the likelihood of identifying relevant studies that may be overlooked by relying on a single database. Our previous work (e.g., ⇡Haßler et al., 2020) suggests that — in education, unlike health — literature databases (e.g., Scopus, ProQuest) only overlap by around 50%, which makes structured approaches across multiple databases necessary. For some topics, the overlap may be as low as 30%. Therefore, it is imperative to search across multiple databases. Moreover, different databases have distinct algorithms and criteria for indexing content, leading to variations in search results. This diversity of sources helps mitigate bias and provides a more balanced view of the literature, as well as validation of our findings through comparison of results.

More specifically, we will conduct a structured automated search of databases with an Application Programming Interface (API) and a structured manual search of high-relevance databases without API using complex queries. We are using digital tools to execute the structured automated search, using existing software development kits (SDKs) to interface with the database APIs. With a fully documented search strategy and automated searches, we can repeat searches at regular intervals, providing future updates beyond the current project, which supports the possibility of a creating a living review beyond the scope of this project. As part of this review, we will use AI-based tools (Scite.ai) to conduct automated forward snowballing and search the reference lists of existing studies. An overview of this approach can be found in ⇡Haßler et al. (2021). In Table 1 below, we have outlined a list of our chosen sources of literature.

We aimed to select the search engines that best facilitate our research objectives. Several search engines were chosen as they include extensive coverage across the relevant topics (i.e., education and technology). Google Scholar is included as its broad scope can allow publications to be discovered that are not found in similar searches in other databases. However, the aforementioned, more established databases will be preferred as they allow searches to be saved and are more transparent in reporting results. Select sources of grey literature will also be included, such as the EEF database of education studies and the Open Development & Education EdTech evidence library, which was created in 2022 alongside the EEF. It contains both meta-analyses and literature reviews relating to school-based EdTech research. A number of evidence libraries found on key NGO websites will be manually searched, namely J-PAL, 3ie, Advanced Education Research & Development Fund (AERDF), and What Works Clearinghouse, selected for their relevance to this project, prominent work in developmental education, and commitment to providing open-source evidence libraries. EEF evaluation reports will be searched for manually, as not all of these can be found in academic databases. As we are not considering theses in our chosen selection of grey literature, we have not included ProQuest.

Table 1. List of chosen sources of literature for the systematic search

Source *(access)*	**Search interface** *Github repository*
OpenAlex *(open access)*	API for structured automated searches https://github.com/OpenDevEd/openalex-sdk OpenDevEd/openalex-cli (github.com)
Scopus *(subscription)*	API for structured automated searches https://github.com/OpenDevEd/scopus-cli
Web of Science *(subscription)*	API for structured automated searches; includes citation trees (cites/cited by) *(A CLI is currently being developed.)*
Scite.ai *(freemium)*	API for citation trees (cited by) OpenDevEd/scite-cli: An unofficial CLI tool that gives access to citation data, scite tallies, related paper metadata and scite reference check. (github.com)
Google Scholar *(free)*	Python -based access for (small scale) structured automated searches; includes citation trees (cited by) OpenDevEd/scholarly-cli (github.com)
British Education Index *(subscription)*	Manual access only
IEEE *(subscription)*	Manual access only

Source *(access)*	**Search interface** *Github repository*
EEF database	Bespoke access
Open Development & Education EdTech evidence library *(open access)*	Manual access only
J-PAL *(open access)*	Manual access only
3ie *(open access)*	Manual access only
AERDF *(open access)*	Manual access only
What Works Clearinghouse *(open access)*	Manual access only

Keyword discovery, search string iteration and conducting initial searches

To identify studies for the review, we have developed an inventory of keywords that we will search for in the title and abstract of studies (↑Education Endowment Foundation & Durham University, 2022;↑EPPI Centre, 2003). The keywords within each category are meant to reflect dimensions within the four levels of EdTech interventions. In Table 2 below, we summarise the categories, sub-categories, and number of keywords within each category.

The initial list of keywords was determined through the interviews conducted, and drawing on existing literature reviews and publications, as well as our prior experience within the research team. Different sets of keywords were trialed using OpenAlex and Google Scholar during the conceptualisation phase to determine appropriate lists of keywords that strike a balance between focus and breadth. In the determination of the keywords we also made decisions about

1. primary keywords: serve as keywords for searches, and
2. secondary keywords: serve to classify and categorise, as well as to eliminate false positives.

The keywords under Item 1 allows us to conduct broad searches producing a large set of search results (~ 200,000). Once the large set of search results has been obtained, basic deduplication (on DOI) will be undertaken, and the results are stored in a database. As we expect a large number of publications to be retrieved, and the data integrity of databases is variable, additional verification of the occurrence of keywords will be undertaken. We will do this by using the keywords under Item 2 to remove false positives from the results to produce a smaller set (~ 10,000), ready for prioritised screening in EPPI Reviewer. Some of the secondary keywords will also be useful to help guide the screening process, such as the list of relevant countries and types of disadvantage that may be mentioned in papers (e.g., Pupil Premium).

To conduct the searches, we will construct search queries using a Command Line Interface (CLI) that systematically incorporate the extensive lists of primary keywords within their respective categories, such as `'technology_category'` and `'education_setting_category'`. Each keyword category is used to construct the full search string, which is passed to the CLI. Within each keyword category, the `'OR'` boolean operator will be used to connect individual keywords. The basic search query is then constructed utilising `'AND'` to join categories.

For example, a search query such as:

```
(technology_category) AND (education_setting_category)
```

would actually search for

```
('educational technology' OR EdTech OR tablet OR laptop OR …)
```

```
AND
('school' OR 'primary education' OR …)
```

This approach ensures a comprehensive search by combining relevant keywords within and between categories.

Our chosen search string (depicted above) will be piloted in May 2024, documenting the number of results for each search. We will manually search through the first 200 results from each database to check relevance and to indicate whether the search is too broad or narrow. Results from these pilot searches will be shared with the EEF to track progress and invite feedback. The search strings may be amended accordingly, for example, we may trial adding in outcomes to help improve the relevance if necessary (either in place of 'education settings' or alongside). All final search strings will be documented and published in the final report for transparency.

Table 2. Overview of keyword inventory

Category	**Description**	**Sub-Categories**	**Number of Keywords**
Technology *(primary keywords)*	The technology under examination in the study	artificial intelligence	247
		audio	
		computer	
		digital personalised learning	
		digital resources	
		EdTech	
		mobile learning	
		online learning	
		social media	
		tablet	
		video	
Education Setting *(primary keywords)*	The educational setting where the study was carried out	early childhood education	52
		primary education	
		primary school	
		secondary education	
		secondary school	
		key stage	
Disadvantage *(secondary keywords)*	The type or types of disadvantage under examination in the study	economic disadvantage	68
		free school meals (FSM)	
		local authority care	
		pupil premium	
Outcomes *(secondary keywords)*	The outcomes under examination in the study	assessment	105
		intermediate outcomes	
		learning outcomes	

Category	**Description**	**Sub-Categories**	**Number of Keywords**
Research Methods *(secondary keywords)*	The research designs and methods of the study	impact evaluation	82
		implementation research	
		literature review	
		meta analysis	
		mixed method	
		monitoring and evaluation	
Geography *(secondary keywords)*	The country or region where the study was carried out	area	79
		country	
Education Terms *(secondary keywords)*	Educational topics under examination in the study	assessment	30
		pedagogy	
		curriculum	
		role	
		materials	
Instructional Domain *(secondary keywords)*	The focus population of the study	generic	87
		literacy	
		mathematics	
		science	
		modern foreign languages	
		Other subjects	

The full keyword inventory can be accessed here.

4.2.2. Stage 2. Classification using topic modelling

We will use Latent Dirichlet Allocation (LDA) – a common machine learning approach for topic modelling (↑Blei & Lafferty, 2009a; ↑Blei & Lafferty, 2009b) – for topic modelling. Topic modelling uses machine learning to identify clusters or groups of similar words within bodies of text — in this case, the publication abstracts. This process will be conducted using readily available libraries in Python, while Zotero will be used for reference management. We plan to explore different LDA characteristics, such as *"number of topics"*, *"number of words per topic"*, *"coherence value"*, to ensure that we identify the best-performing characteristics to create the most relevant topics for our study. Examples of relevant topics could be 'digital learning platforms' or 'online assessment and feedback'. Once the research team has created a provisional topics list, we will consult with the EEF to identify any topics that may be missing and ensure that the topics modelled are relevant to teaching and learning practices in England (as the focus of this review). In doing this, we want to identify the optimal number of topics to capture relevant studies.

Researchers will review the emerging topics, which will lead to a refinement of the precise LDA parameters utilised. The review team will repeatedly pilot this process with the sample until satisfied that a sufficiently broad and relevant set of topics has been captured. The LDA will indicate how publications are characterised, which is helpful for the evidence map; the LDA may also indicate promising areas with sufficient publications to conduct meta-analysis.

4.2.3. Stage 3. Eligibility screening of titles, abstracts, keywords

The long list of studies will be uploaded to a specialist review software, EPPI-reviewer (Version 6), for manual screening, document management, data extraction, and data analysis.

As there may still be duplicate results after the initial deduplication based on DOI (especially as not all grey literature has a DOI), we will use the inbuilt deduplication tool within EPPI. We will use the automated marking of duplicates, setting the threshold at 85% similarity. For remaining papers with very high similarity scores, reviewers will manually check the title, abstracts, and meta-data, before deciding whether the papers are duplicates.

Two reviewers will review the abstracts of studies on the long list, further coding studies into 'high', 'medium', and 'low' relevance categories. We have chosen this approach rather than the traditional include / exclude as we feel that presenting three categories allows us to capture uncertainty better, which is critical in the somewhat messy and ill-defined field of EdTech. This more nuanced approach to study selection will allow us to have further conversations as a research team where a paper's relevance is unclear. While this may lead to higher levels of moderation than a standard review, we would prefer to undertake this than risk losing lots of potentially relevant papers.

For this review, relevance refers to how well a publication aligns with our focus on the use of EdTech to support the academic attainment of disadvantaged pupils, based on the information that is provided in the abstract. The criteria can be classified such that:

- High (H): clearly satisfactory
  Publications that clearly and directly address the use of EdTech to support academic attainment. For example, an RCT that specifically investigates the effectiveness of an EdTech intervention in improving the academic performance of students from a low-income background in the UK.
- Medium (M): unclear or contentious
  Publications where the relevance to our focus is unclear or open to interpretation. For example, a quasi-experimental study examining the impact of an EdTech tool on student engagement across different socio-economic backgrounds without explicitly stating whether the study measures its impact on academic achievement.
- Low (L): clearly unsatisfactory
  Publications that do not address our focus adequately or are unrelated to the use of EdTech for supporting academic achievement. For example, a qualitative case study exploring the opinions of technology use in training student nurses. (This example was selected as studies related to health education often appear in educational searches as false positives due to similar keywords, e.g., 'student'.)

The coding of both reviewers will be compared to explore similarities and conflicts in relevance ratings. The coding tool for this stage can be found in Annex 4 and is based on the inclusion and exclusion criteria below (Table 3).

- Where reviewers agree articles are of high relevance (H / H), the articles will be included in the study.
- Where one reviewer codes an article as high relevance and the other reviewer codes the same paper as medium relevance (M / H or H / M), the article will be reviewed in a meeting with both reviewers to reach intercoder agreement. Where no agreement is reached, a third (senior) reviewer will review the paper and make the final decision, taking feedback from the two reviewers into consideration.
- Where reviewers agree articles are of medium relevance (M / M), a third (senior) reviewer will screen the title and abstract. This reviewer may reclassify the articles as low relevance if the focus (e.g., technology) is not adequately represented or the abstract does not make it clear enough that the paper could be relevant to the review. They may also reclassify to high if they feel it is clearly relevant. All remaining medium-relevance articles will be screened on full-text to avoid missing relevant studies.
- Where one reviewer codes an article as low relevance and the other reviewer codes the same paper as medium or high relevance (L / M or L / H), the article will be

reviewed in a meeting to reach intercoder agreement. Where no agreement is reached, a third (senior) reviewer will review the paper and make the final decision, taking feedback from the two reviewers into consideration.

- Where reviewers agree articles are of low relevance (L / L), the articles will be excluded from the study.

At this stage, low-relevance studies will be excluded and medium- and high-relevance studies will be included in a shortlist for Stage 4: full-text screening.

Additional automated screening: Given our significant interest in how AI can be utilised to enhance research processes, in parallel, we will also use several LLMs (e.g., ChatGPT, Llama, Mixtral) to automatically screen and code studies in the long list based on available metadata (e.g., title, abstract, keywords). This will be an additional research step that does not impact the current review, but can potentially serve to increase knowledge of how we can utilise AI tools in the search and screening stages of a review.

The screening tool used for the manual review process will be fed to the AI review tool, and questions trialled on a small number of publications (up to 50 per iteration). Researchers will manually evaluate the AI's responses to the screening questions against the abstracts to determine the output validity and reliability. Questions will be adjusted as needed based on AI responses until the research team agree that AI responses are accurate over several trials.

Subsequently, the outcomes of manual and automated screening will be compared and moderated, ensuring consistency and accuracy in the selection process. This will include identifying areas of agreement and discrepancy between the decisions made by human reviewers, and those generated by the AI-based screening tool. Discrepancies will be highlighted and discussed within the team before adjusting the input (screening questions) to optimise the performance of the AI tool, and improve alignment with human judgement to avoid the model rejecting potentially relevant studies. This dual approach, combining manual and AI screening capabilities, aims to provide evidence to streamline the screening processes, while enhancing efficiency and reducing potential bias, in the context of future research (forthcoming conference contribution). For the purposes of the present study, the human screening is utilised for decision making about inclusion and exclusion.

To be included in this review, studies will need to meet the inclusion criteria listed in the table below.

Table 3. Inclusion and exclusion criteria

Category	**Inclusion criteria**	**Exclusion criteria**	**Justification**
Date of publication	Papers published since 2011	Papers published before 2011	We will restrict our search to studies published after 2011 as: ● this field is rapidly evolving, with new technologies and digital approaches frequently emerging in the classroom (↑Lewin et al., 2019) ● relatively limited rigorous evaluation data has been gathered on the impact of EdTech interventions in the classroom (↑Haßler et al., 2021). ● 2011 is the year before the publication of EEF's first systematic review on EdTech (with the second published in 2019). As the report was published in November 2012, this should leave no gap between the two searches (as the 2012 review included papers from 2012).
Publication type (primary focus)	Peer-reviewed journal articles Working papers (technical reports, policy papers, conference papers) Select additional grey literature to include: EEF evaluation reports, and Research found on the:	Books Book chapters Grey literature (other than those mentioned in inclusion criteria) Extended abstracts Theses	We will include studies reported in peer-reviewed journals as these studies are assessed for rigour, quality, and originality. However, relevant studies from peer-reviewed journals will be manually reviewed for rigour and quality before determining final inclusion.

Category	Inclusion criteria	Exclusion criteria	Justification
	EEF database Open Development & Education EdTech evidence library EdTech Hub evidence library J-PAL website 3ie website AERDF website What Works Clearinghouse		We will include working papers, as contemporary uses of educational technologies have not been the subject of peer-reviewed evaluation (↑Cheung & Slavin, 2013; ↑Lewin et al., 2019). Similarly, these will be manually reviewed for rigour and quality to determine inclusion. We will also include select grey literature from sources identified as potentially relevant.
Research design (primary focus)	Studies with rigorous causal inference strategies, including randomised controlled trials and quasi-experimental methods (e.g., instrumental variables, differences-in-differences, fixed effects, regression discontinuity, propensity score matching) Mixed-methods studies that include a quantitative element with a rigorous casual inference strategy.	Studies without rigorous causal inference strategies, such as: Qualitative studies Cross-sectional studies lacking longitudinal / experimental elements. Non-experimental designs	We will restrict our search to studies with rigorous causal inference strategies, as when undertaken robustly, these offer the potential to provide unbiased and internally valid estimates on causal impact.

Category	Inclusion criteria	Exclusion criteria	Justification
Language of publication	English Exploratory: French, Spanish, Chinese	Publications in languages other than English, French, Spanish, or Chinese.	We will *primarily* focus on studies in English. We will also conduct additional, exploratory searches in French, Spanish, and Chinese using OpenAlex to explore the extent of the limitation that a sole focus on English presents.
Geographic focus	Countries with 'high technological readiness' on the United Nations Conference on Trade and Development's 2023 Technological Readiness Index (see Annex 5 for the list of specific countries)	Countries not identified as having 'high technological readiness'.	We will focus on countries with 'high technological readiness' as they have similar digital landscapes as England and the wider United Kingdom. The United Nations Conference on Trade and Development used the following indicators to identify countries with 'high' technological readiness: ● **ICT deployment:** internet users (as a percentage of the population) and mean internet download speed (in megabits per second) ● **Skills:** expected years of schooling and high-skill employment (as a percentage of the working population) ● **Research and development activity:** number of scientific publications on frontier technologies and number of patents filed on frontier technologies ● **Industry activity:** high-technology manufacturing exports (as a percentage of total

Category	**Inclusion criteria**	**Exclusion criteria**	**Justification**
			merchandise trade) and digitally deliverable services exports (as a percentage of total service trade) ● **Access to finance:** domestic credit to private sector (as a percentage of GDP)
Population (P)	Students in formal mainstream education in KS1–KS5 (ages 5–18). Disadvantaged students in formal education in KS1–KS5. For this study, we will primarily focus on the UK Government's definition of disadvantage: ● eligible for free school meals*; ● have been recorded as eligible for free school meals in the past 6 years; ● under the care of a local authority, or previously under the care of a local authority.	● Children who are too young to attend formal schooling (<5) ● Students in higher education (HE) ● Adults ● Out-of-school children (any children not in formal education) ● Home-schooled children ● Children with Special Educational Needs or Difficulties (SEND) who attend a specialist setting.	We will search for studies that include *all* pupils in formal education in KS1–KS5 to avoid conducting a search that is too narrow, or missing studies that investigate disadvantage but as a subgroup within a wider analysis. Including wider school-based EdTech studies may also help us identify mechanisms that could be transferable for disadvantaged pupils. Our prioritised focus is on studies that study disadvantaged students in formal education in KS1–KS5, as we intend to investigate how the mechanisms and building blocks of EdTech interventions interact with disadvantage. *As 'free school meals' is an indicator of disadvantage that is specific to the UK, we will also consider similar / equivalent initiatives that may be found in other included countries. A (non-exhaustive) list of these is provided in Annex 5.
Intervention (I)	Interventions using any hardware, software, other digital	● Interventions not involving EdTech	We opted to exclude any studies focusing on non-student outcomes (e.g., teacher attendance) as we

Category	Inclusion criteria	Exclusion criteria	Justification
	approaches, or a combination of these components alongside a defined set of activities and materials aimed at implementing technology that is selected by schools with the explicit and primary goal of improving student academic attainment / achievement. This includes EdTech implemented by schools but used outside of school (e.g., EdTech approaches to supporting homework).	● Interventions focused on hardware only ● Interventions implemented by parents / carers ● Interventions focused on teacher professional development (even if related to EdTech) ● Interventions not used with the primary goal of improving academic achievement.	are going to examine the mechanisms and building blocks that underpin EdTech interventions which improve student attainment. As noted above, we adopted a deliberately broad definition of EdTech as evidence shows that choices on the use of technology in the classroom have a larger impact on learning than the choice of technology (↑Higgins et al., 2012; ↑Lewin et al., 2019; ↑Verbruggen et al., 2021; ↑Walker et al., 2023). Separately, we have not focused on any specific pedagogies to avoid making assumptions about the mechanisms and building blocks that underpin successful interventions.
Comparison (C)	No intervention, non-EdTech intervention, or waitlist intervention for a control group Another EdTech intervention as a comparison group	No control or comparison group	
Outcomes (O)	Studies using tests to quantitatively measure student attainment or academic achievement in any curriculum subjects	Observational assessments Qualitative assessments	We will focus on outcomes in all curriculum subjects to widen the scope of EdTech reviews beyond specific domains such as English and maths (e.g.,↑Abrami et al., 2020; ↑Verbruggen et al., 2021). As such, we have avoided specifying the type of assessment (e.g., EGMA / EGRA) to avoid limiting our subject focus.

Category	Inclusion criteria	Exclusion criteria	Justification
			We have excluded observational and qualitative assessments to allow rigorous and consistent comparison across schools and studies.

4.2.4. Stage 4. Initial full-text manual screening

After the title and abstract screening, the full texts will be retrieved for studies that appear to meet the criteria (high relevance) and those that are borderline (likely due to insufficient information; medium relevance). These papers will be screened based on the full-text of the article, using the inclusion criteria and the same 'high / medium / low' relevance rating system.

For this stage, partial double screening will be used. This will consist of double screening 20% of records and then switching to single screening if 95% agreement has been achieved. If not, then double screening will continue until 95% has been achieved. At the end of the full-text screening process, papers considered to be of 'medium relevance' will be double-screened, and the same agreement process will be applied as used for the title and abstract screening. As before, the articles may be reclassified as having low relevance if the focus (e.g., technology, disadvantage) is not adequately represented. Any remaining medium-relevance articles may be included in the evidence map, but with the caveat that we have identified them as not highly relevant. The results of this process will be documented in the PRISMA flow chart found in Annex 3.

4.2.5. Stage 5. Evidence mapping

To address sub-research Question 6, "What are the key characteristics of school-based EdTech intervention studies implemented in countries with 'high technological readiness'?", we will create a systematic evidence map to provide a structured overview of existing research on EdTech interventions within this context. We aim to keep the focus of this map relatively broad (as is the nature of evidence maps), but will also code for elements specific to this review (such as disadvantage). All studies of high and medium relevance after the full-text screening will be coded for the following:

- Publication year
- Age group / school stage (e.g., Preschool, primary, secondary, college)
- Country of intervention
- Technology tools used:
    - hardware
    - software
- Academic subject(s) targeted
- Type of intervention, including
    - Name of intervention (if branded entity in its own right)
- Modes of delivery, including
    - Location of learning activities (e.g., classroom, playground, home)
- Focus on disadvantage? (Yes — full paper, partial focus, no)
    - Type of disadvantage explored (e.g., free school meals, low socioeconomic status, SEND)

- Methodology (e.g., RCT, QED, mixed methods)
- Primary outcome
- Test type (e.g., standardised test, test developed for purpose by researcher / teacher)
- Relevance to research question (high or medium — as decided at full-text screening)

Data extraction for the map will be piloted to check whether any clarification or refinement is necessary. Two reviewers will conduct data extraction on a number of randomly selected studies (20% of the total papers). As with all other stages, cases of disagreement will be resolved by discussion, with input from a third, senior reviewer if the disagreement does not get resolved. This will highlight areas where clarification might be needed for coding, so any points of confusion will be addressed and amended. Once agreement has been reached, full independent data extraction will commence. If concerns arise during the coding process, the research team will meet to decide whether refinements to the coding are necessary.

The resulting evidence map will be presented in a user-friendly tabular format to maximise use and accessibility.

4.2.6 Stage 6. Full-text quality appraisal

After producing the evidence map, the high- and medium-relevance studies that passed the full-text screening will be assessed for quality. For the purpose of this review, quality refers to a paper's methodology and rigour. We will use an adapted version of an existing quality appraisal tool that has been used in education research, the Mixed Methods Appraisal Tool (↑Hong et al., 2018). All sections that are relevant for our inclusion criteria will be utilised, so we will not use Section 1 (qualitative research) or Section 4 (quantitative descriptive research), as these studies would not have passed our full-text screening. For quantitative RCTs, we will not utilise the question about blinding, as this is often unfeasible in education research. The finalised tool we will use is shown in Table 4. The instructions provided alongside the original tool to guide responses will be utilised in full for the remaining questions.

The review team will code 10 randomly selected interventions in a group quality appraisal meeting, and discuss any discrepancies in the meeting. Full agreement will need to be achieved before embarking on independent coding. If concerns are raised about the tool, then the team will decide whether further refinement to the tool is necessary to clarify the differences in interpretation. Reviewers will independently code all the remaining studies.

If multiple papers report the same study, then we will retrieve all versions and assign one as the primary document, which would likely be the most detailed and / or most recent journal article. Linked papers will be reviewed when there is information not reported in the primary document to see if they contain the additional information.

Authors of the tool do not recommend calculating a score based on the answers to the questions listed in Table 4 below. We also do not think that a score is suitable, as some questions may hold more weight than others for study quality. As a result, we will conduct a quality appraisal meeting within the research team to discuss the papers and corresponding answers on the quality appraisal tool, and decide collaboratively whether a paper is high, medium or low quality. At this stage, for the final review, data will only be extracted from high-quality papers.

We acknowledge that a high number of studies may need to be quality assessed, so if over 150 studies are identified following the full-text screening, this process may be adapted in consultation with the EEF.

This quality assessment stage will ideally result in a final set of high-relevance, high-quality studies being selected for data extraction in EPPI Reviewer. If only a few high-quality studies are found, we would discuss with the EEF an alternative plan to also data extract the medium-quality papers.

Table 4. Quality appraisal tool

Category	Questions	Responses
Screening (for all papers)	Are there clear research questions?	Yes / No / Can't tell
	Does the collected data allow the researchers to address the research questions?	Yes / No / Can't tell
Quantitative RCT	Is randomisation appropriately performed?	Yes / No / Can't tell
	Are the groups comparable at baseline?	Yes / No / Can't tell
	Is there complete outcome data?	Yes / No / Can't tell
	Did the participants adhere to the assigned intervention?	Yes / No / Can't tell
Quantitative non-randomised	Are the participants representative of the target population?	Yes / No / Can't tell
	Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes / No / Can't tell
	Is there complete outcome data?	Yes / No / Can't tell
	Are the confounders accounted for in the design	Yes / No / Can't tell

	and analysis?	
	During the study period, is the intervention administered as intended?	Yes / No / Can't tell
Mixed methods	Is there an adequate rationale for using a mixed-methods design to address the research question?	Yes / No / Can't tell
	Are the different components of the study effectively integrated to answer the research question?	Yes / No / Can't tell
	Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	Yes / No / Can't tell
	Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	Yes / No / Can't tell
	Do the different components of the study adhere to the quality criteria of each tradition of the methods involved? *(Note: for this question we would ensure that reviewers had access to the qualitative quality criteria so they could assess appropriately)*	Yes / No / Can't tell

4.2.7. Stage 6. Data extraction and management

The research team will use EPPI Reviewer to upload, manage and extract data from all records. To this end, we will use the EEF's existing main data extraction (MDE) and effect size data extraction (ESDE) tools in their entirety, as a framework to code, highlight, and extract relevant information (↑Education Endowment Foundation & Durham University, 2022). An additional supplementary tool, created by the research team, will be used to capture details relevant to the current research that are not fully covered by the EEF tools. The supplementary tool consists of six high-level steps, as summarised in Table 5 below. The supplementary tool is available ↑here.

Table 5. Data extraction procedure

#	Step	Description
1	Disadvantage	Reviewers will code a little more for disadvantage to cover areas not mentioned in the MDE tool (and capture equivalent terms from countries outside the UK).
2	Educational technology and pedagogy	Reviewers will identify which technology hardware, software, and approaches were used.
3	Mechanisms and barriers	Reviewers will inductively extract data on mechanisms utilised in the intervention and discussed in the paper. Reviewers will extract data on any barriers to the EdTech intervention identified by the authors.
4	Intermediate outcomes	Reviewers will extract data on any relevant intermediate outcomes that have been tested as part of the intervention (e.g., pupil attitudes, engagement, motivation, participation or attendance).
5	Study transparency	Reviewers will extract data on any methodological limitations identified by the authors.
6	Additional information	Reviewers will provide any additional information from the study that may be relevant and has not already been covered, and identify if

		they see a reason to contact the study authors for more details / missing information.

To test the tool, two reviewers will code extracted data from 5 randomly selected studies that have passed the detailed manual quality screening (see Section 4.2.6). If and when discrepancies emerge, the reviewers will reach an agreement either through discussion or with the support of a third reviewer. Feedback from this process will be collected and used to refine the coding tool (e.g., adding any additional hardware/software options, making definitions or examples more clear.)

Afterwards, two reviewers will use the refined tool to double-code a random sample of 20% of the complete set of records. As mentioned above, the reviewers' outputs will be compared to assess inter-coder reliability. If the level of coding agreement is below 80%, the research team will explore the possibility of conducting additional double-coding or making further revisions to the coding tool. If no concerns emerge, reviewers will proceed with single coding for the remainder of the records.

4.2.8. Effect size calculation

Effect sizes will be calculated in EPPI-reviewer, using the EEF's existing effect size data extraction tool. Meta-analysis is typically a two-stage process. In the first stage, a summary statistic is calculated for each study, to describe the observed intervention effect in the same way for every study. For example, the summary statistic may be a risk ratio if the data are dichotomous or categorical, or a difference between means (standard mean differences like Cohen's d and hedges g) if the data are continuous. Given our meta-analysis focuses on student achievement measured by continuous test scores, we will employ Cohen's d as the primary effect size metric. This statistic is typically calculated using means and standard deviations (or standard errors and confidence intervals) available from each study. If these primary statistics are unavailable, we will utilise alternative effect size estimators, prioritising t- or F-statistics over p-values due to their limitations. Formulas and web resource for calculating the various standard mean differences and other effect sizes are given below[1]. In the second stage, pooled effect sizes will be calculated (see more information in Section 5.1).

4.2.9. Unit of analysis issues

While most cluster-randomised trials adjust for nesting in the data, some studies fail to report appropriate analyses. Sometimes analysis is conducted as if the randomisation was

[1] Formulas and Software link:
https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html

performed on the individuals rather than the clusters. If this is the situation, approximately correct analyses may be performed if the following information can be extracted:

- The number of clusters (or groups) randomised to each intervention group and the total number of participants in the study; or the average (mean) size of each cluster.
- The outcome data ignoring the cluster design for the total number of individuals (e.g. the number or proportion of individuals with events, or means and standard deviations for continuous data); and an estimate of the intra-cluster (or intra-class) correlation coefficient (ICC).
- The ICC is an estimate of the relative variability within and between clusters. Alternatively, it describes the ‘similarity’ of individuals within the same cluster.

In instances where effect size data is absent from a study, several strategies can be employed to address this shortcoming. Our preferred course of action will be contacting the study's corresponding author with a direct request for the missing effect size information. This approach prioritises the inclusion of the study's data while potentially improving the overall comprehensiveness of the meta-analysis.

If such efforts to acquire the missing data prove unsuccessful, alternative strategies will be explored. One approach involves attempting to calculate the effect size from the information provided within the study itself. This might include utilising statistics such as beta coefficients and standard errors, provided they are reported. However, the feasibility and accuracy of this approach depend on the specific information available and the chosen effect size metric.

In rarer circumstances, the research may have associated replication materials available. These materials might contain the necessary data to independently calculate the effect size, potentially salvaging the study's inclusion in the meta-analysis.

As a final resort, if none of the aforementioned strategies yield a solution, study exclusion from the meta-analysis might be necessary. This course of action would be carefully considered and justified, as it reduces the available data and potentially weakens the generalisability of the meta-analytic conclusions.

This brings us to the broader question of how to address issues related to missing data for any effect size calculations. There are many potential sources of missing data in a systematic review or meta-analysis. For example, full papers may be missing for studies identified in the initial search might be missing online and hence are not incorporated in the review process. In such cases, we will first try to contact the authors to get hold of such papers and if not available, a list of such missing studies will be maintained.

Missing outcome data within a study will also be documented. We will prioritise utilising raw data or imputing missing values based on available data whenever possible. For instance, summary data for an outcome might be absent, but combining results from relevant subgroups could yield a complete value for the overall sample. This approach will be favoured over discarding the effect size due to missing data. A meticulous record of all calculations and imputations will be maintained. Subsequently, a sensitivity analysis will be conducted to assess the potential impact of these procedures on the observed effects (e.g., bias). Finally, participant-level characteristics or values might be missing from retrieved summary data. In such instances, imputation techniques outlined by Akl et al. (2015) will be employed, as feasible.

4.2.10. Content analysis

A primary objective of this research is to identify the core mechanisms of EdTech interventions that lead to improved attainment outcomes for students. By this stage of the research, we will have created a working ‘mechanisms map’ from the experiences and expertise of education practitioners and existing literature. Mechanisms will then be inductively extracted from the papers in this review during the data extraction phase, along with corresponding definitions / explanations around their use and implementation. We have chosen not to deductively code using the existing mechanisms map to avoid restricting the search to those already identified, and our interest in identifying mechanisms that have been used in interventions that led to improved attainment outcomes.

The extracted mechanisms and surrounding explanation will be imported into Atlas.ti to be qualitatively analysed. At this stage, it is likely that we will choose a form of content analysis to analyse the data on mechanisms (e.g., deductive content analysis using our existing ‘building blocks’), but this will be confirmed once we have firm examples to understand the scope of explanation provided within studies. We anticipate that the data will not be rich enough for a thematic analysis, but would consider this as an alternative method should we find that the dataset is rich and suitable. The final, chosen methodology will be included in an analytical addendum, which will be published before analysis takes place.

This analysis will be used to inform and update the mechanisms map created in Phase 1, which will be one of the artefacts presented in Phase 3. Depending on the number and quality of the remaining studies, it may also be used to inform some of the subgroup analyses that we are interested in conducting for the meta-analysis concerning some of the more prominent mechanisms.

4.3. Phase 3: Structured Community Review

As previously identified, we strongly believe that this review should be informed by the lived experience of education practitioners to address the balance between what is considered effective in education research and what is actually reflected in classroom practice. Therefore, in Phase 3, we will conduct a structured community review to consult education practitioners with the results of the Phase 2 systematic review, allowing us to test and further develop our insights on effective EdTech use. This approach will provide grounding and assessment of feasibility to this mixed-methods research.

A structured community review entails a structured process to engage experts and practitioners in the review of a draft report (e.g., a systematic literature review). This process involves semi-structured interviews, focus groups, and online surveys. This approach shares aspects with realist reviews. ↑Pawson et al. (2005) note:

> *"traditional methods of review [in the social sciences] focus on measuring and reporting on programme effectiveness, often find that the evidence is mixed or conflicting, and provide little or no clue as to why the intervention worked or did not work when applied in different contexts or circumstances, deployed by different stakeholders, or used for different purposes."* (p. 21)

This issue is particularly acute in education research. For instance, several reviews have concluded that 'the linkage [between technology and attainment] may not be a simple causal one nor necessarily a simple linear association' (↑Higgins et al., 2012: p.7; ↑Lewin et al., 2019). Relatedly, the best quantitative studies:

> *"are still likely to lead to erroneous conclusions because they cannot distinguish the active ingredients of rigorously evaluated interventions from the causally redundant components."* (↑Sims & Fletcher-Wood, 2020: p.3)

For our structured community review, we will ask education practitioners (consisting of teachers, teaching assistants, specialist teachers, and educational psychologists) in England to share evidence-based insights and practice-based critiques of our conclusions. In particular, we will use semi-structured interviews and an online survey to solicit qualitative feedback on the mechanisms map and select results from the meta-analysis. The final artefacts will depend on the analyses conducted (confirmed in a later addendum). In soliciting feedback on our artefacts, we aim to validate our understanding of:

- the 'building blocks' (Level 2) and 'mechanisms' (Level 1) that improve attainment outcomes for pupils;
- the intermediate outcomes associated with increased pupil attainment within each theory of change;
- different routes to impact, associated with different mechanisms;
- the relevance of different theories of change to socially and economically disadvantaged pupils, which is the primary focus of our review.

For our sample, we aim to gather a diverse sample of practitioners, as we believe all practice-based insights will be valuable for the development of our insights and understanding. However, as the specific focus is on disadvantage, we aim to ensure that we obtain at least a sample of practitioners serving disadvantaged students to obtain the specific lived experiences of working with this population of children. Participants for the interviews will be identified through existing contacts and networks; two members of the research team are teachers and one is an Educational Psychologist. Where possible, we will aim to reinterview the teachers identified in Phase 1 of this study to share our findings following initial discussions. The community review survey will be advertised on the EEF and Open Development & Education social network platforms, similarly to the survey from Phase 1.

Insights from this community structured review will further be incorporated into the mechanisms map and reported alongside the main research synthesis (meta-analysis and content analysis) in the final report.

5. Data synthesis

Following the review steps outlined in Section 4.2, a meta-analysis of the systematically collected, high-quality literature is planned to answer sub-research Question 6 “What is the impact of school-based EdTech interventions on attainment outcomes in countries with ‘high technological readiness’?” Subsequent analyses (to be decided at a later point based on the evidence base) will be conducted to help answer the more granular, but arguably more important, research questions around the impact of mechanisms and how the overall findings apply to disadvantaged children.

This plan details the proposed higher level methods, but development of the full analysis plan is ongoing and will largely depend on the results of the searches and thus cannot be pre-specified at this point. Once the heterogeneity has been calculated, the full statistical analysis plan will be confirmed and published in an addendum prior to the analysis being undertaken.

For the meta-analysis, the commonly used inverse-variance method will be used. The inverse-variance method is so named because the weight given to each study is chosen to be the inverse of the variance of the effect estimate (i.e., 1 over the square of its standard error).

Thus, larger studies, which have smaller standard errors, are given more weight than smaller studies, which have larger standard errors. This choice of weights minimises the imprecision (uncertainty) of the pooled effect estimate. This can be done using two methods: the fixed-effect and the random effect method. This meta-analysis will employ a random-effects model. This choice is justified by the inherent heterogeneity in educational research and the desire for generalisable findings. Random-effects models are generally preferred for meta-analyses in education, particularly when dealing with student test scores across various subjects. This approach acknowledges the variability in true effects between studies and provides more conservative estimates that can be applied to a broader range of educational contexts.

Meta-analysis is a powerful tool for synthesising research findings when specific criteria are met. In order to determine whether the approach is feasible and meaningful, we will: first ensure that the studies included will investigate the same constructs and relationships, i.e. to ensure they address the same research questions. Second, the studies need to report their findings in a statistically comparable format. This could be effect sizes, correlation coefficients, odds ratios, or other metrics that can be quantitatively combined. Finally, for the studies to be considered *"comparable"* in the context of the meta-analysis, they should share key characteristics relevant to the research question. These characteristics include:

- **Objective:** Is the focus on the overall effect of EdTech on student attainment or exploring variability in effects across different studies/contexts/implementations?

- **Population:** Do the studies involve similar participant demographics (i.e. children/adolescents who attend school or higher education)?
- **Study design**: Do all studies construct a counterfactual? Do all studies clearly define their control/comparison groups? Do the studies incorporate the same mechanisms within their intervention?
- **Individual characteristics:** Do the studies involve participants with similar characteristics relevant to the research question (e.g., age cohort, grade in which student is studying, and so on)?

Based on prior knowledge of the research field, we do not anticipate the meta-analysis to be deemed unfeasible. However, should the meta-analysis be deemed unfeasible or not meaningful, we will discuss this as a team with the EEF to communicate our conclusions. If they agree that the approach is now unfeasible, we will conduct a systematic literature review using the existing pool of high-quality studies.

We acknowledge the importance of minimising the inclusion of studies with multiple effect size calculations derived from the same sample. This practice can inflate the weight of a single study and introduce bias into the analysis. If complete exclusion of such studies proves infeasible, Robust Variance Estimation (RVE) will be employed to adjust for the dependence between effect sizes. RVE accounts for this dependence by modifying the standard error of the pooled effect size, leading to more accurate confidence intervals. However, it is crucial to acknowledge a key limitation of RVE: its effectiveness relies on a sufficiently large number of studies in the analysis (↑Pustejovsky & Tipton, 2021). The meta-analysis will utilise the Correlated and Hierarchical Effects (CHE) model within the RVE framework. This model acknowledges the reality that meta-analytic data often exhibits a mixed structure, where both correlated and hierarchical effects are likely to be present (↑Pustejovsky & Tipton, 2021). Specifically, the analysis will implement the random-effects model with Knapp-Hartung adjustment as a robust variance estimation technique within the CHE framework. This approach addresses dependence between effect sizes while acknowledging the presence of random variation across studies.

To answer sub-research Question 3 “What intermediate outcomes are associated with improved pupil attainment in EdTech interventions, and where do they fit within the identified mechanisms?”, reviewers will data extract for any intermediate outcomes that have been tested as part of the intervention, and identified as relevant for the current review. These intermediate outcomes are pupil attitudes, engagement, motivation, participation, and attendance. As with the primary meta-analysis calculation, if there are enough papers that are high-quality and comparable (according to the characteristics criteria above) for a specific intermediate outcome, such as engagement, we will conduct a meta-analysis on this subset of papers to determine the association between engagement and pupil attainment in the context of EdTech interventions. In papers where the mechanisms are explicitly linked to the intermediate outcomes of interest (e.g., the utilisation of an animal character within an app led to increased engagement in the classroom), this will be incorporated into our existing mechanisms map with the 'if… so… then' structure to identify causal links between mechanisms and the outcomes. Similarly, if an intermediate outcome is explicitly stated to impact a mechanism, this will also be mapped. We note that the primary outcomes will be extracted first; where intermediate outcomes are available, the type of intermediate outcome is noted, and publication is marked for further extraction. The feasibility of this planned analysis will be determined and finalised in a future addendum.

We will be presenting a mixed-methods synthesis of the meta-anlysis results, whereby the qualitative data on mechanisms from the content analysis, plus further insights from practitioners, will be used to provide explanation, context, and depth to the quantitative data (e.g., subgroup analyses effect sizes for specific mechanisms). Further details will be provided once we have established the full analysis plan and can identify what is feasible.

5.1. Pooled effect size calculation

Pooled effect size calculations are made by estimating the combined intervention effect by calculating the weighted average of the intervention effects estimated in the individual studies.

$$weighted\ average = \frac{\sum (estimate * weight)}{\sum weights} = \frac{\sum Y_i W_i}{\sum W_i}$$

Where Y_i is the intervention effect estimated in the i^{th} study, W_i is the weight given to the i^{th} study, and the summation is across all studies. Note that if all the weights are the same, then the weighted average is equal to the mean intervention effect. The bigger the weight given to the i^{th} study, the more it will contribute to the weighted average.

Effect size in its broad sense can refer to both unstandardised and standardised effect measures for both categorical and continuous outcomes. For continuous outcomes, raw mean difference and standardised mean difference (*SMD*) are the most frequently used for unstandardised and standardised effect measures, respectively. *SMD*, also known as the d family estimator, is used to measure the mean difference between groups in terms of standard deviation (*SD*). While various methods exist for calculating standardised mean differences (*SMDs*), this meta-analysis will utilise the post-test (endpoint) score approach due to the continuous nature of our outcome variable, student achievement scores.

Post-test (endpoint) score approach (using the post-test mean difference as the numerator):

$$SMD = \frac{M_{I,post} - M_{C,post}}{SD_{pool,post}}$$

This is the classic Cohen's d (or Hedge's g when corrected for small sample size), obtained by dividing post-test score mean difference by the pooled post-test score *SD*. It is straightforward and consistent with the classic two independent group t-test. It may well be the case that a paper (or, indeed, several papers concerning the same intervention) measure more than one outcome, such as mathematics and literacy outcomes. In such cases, considering the degree of independence of the outcomes, we will make a decision whether one or more outcomes will be taken forward for the analysis.

However, some researchers argue that sometimes it is implausible to assume that the post-test scores of two groups have equal variance. This is because various individual treatment responses can inflate SD, but the inflation may be disproportionate in the treatment and control group. This is $Glass's\ \Delta$,

$$SMD = \frac{M_{I,post} - M_{C,post}}{SD_{pool,pre}}$$

The denominator is the untreated SD (which can be the control group SD or pooled pre-test SD; as the former only uses information from one group, the precision would be compromised, thus pooled pre-test SD is usually preferred). It has an added advantage, we do not need to assume equal variance of post-test scores, and may be more stable and consistent across trials or studies.

Following the completion of the primary meta-analysis, we will conduct subgroup analyses to explore whether the effects of an EdTech intervention vary across different groups. For this review, primary areas of interest for subgroup analyses would be disadvantaged children, and exploring the impact of different mechanisms identified in the literature (e.g., if multiple papers identify using the same mechanism, we could conduct a subgroup analysis). However, to determine what is feasible, we first need to conduct the heterogeneity analysis that will help us identify potential sources of variability. The planned heterogeneity analysis is outlined below. A full analysis plan will then be published as an addendum to this protocol outlining which subgroup analyses will be undertaken.

5.2. Investigation of heterogeneity

Three statistics will be calculated to assess the presence of statistical heterogeneity: Q-statistic, I^2 and τ^2. It is essential to consider the extent to which the results of studies are consistent with each other. If confidence intervals for the results of individual studies (generally depicted graphically using horizontal lines) have poor overlap, this would generally indicate the presence of statistical heterogeneity. Here, Homogenous Statistic Q or Chi2 would help us to assess whether observed differences in results are compatible with chance alone. A low P value (or a large Q or Chi2 statistic relative to its degree of freedom) would provide evidence of heterogeneity of intervention effects (variation in effect estimates beyond chance). Methods can also be developed for quantifying inconsistency across studies that move the focus away from testing whether heterogeneity is present to assessing its impact on the meta-analysis. A useful statistic for quantifying inconsistency is:

$$I^2 = \left(\frac{Q-df}{Q}\right) * 100\%$$

I^2 describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance). Various thresholds of I^2 will be used to conclude towards presence of heterogeneity, namely: 0% to 40%: might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may represent substantial heterogeneity; and 75% to 100%: considerable heterogeneity. τ^2 will be calculated to derive the between-study variance in our meta analysis.

5.3. Sensitivity analysis

We will investigate the robustness of the findings by conducting a series of sensitivity analyses. These analyses will explore how the overall effect size and conclusions might change under different assumptions or conditions. The aim is to assess the potential influence of various factors on the results and ensure that the overall findings are robust.

Implementation issues:

- Compliance: The analysis will exclude studies where a high proportion of teachers reported not effectively integrating the EdTech intervention into their teaching practices. This could include instances where teachers used the technology superficially, did not use it for its intended purpose, or significantly deviated from the intervention's guidelines.
- Attrition: Studies with a high number of participants who dropped out of the study or have missing data points will be excluded. The criteria for high attrition will be established based on existing guidelines, such as the What Works Clearinghouse (WWC) standards.
- Sample size: Separate analyses will be conducted excluding studies with very small sample sizes (e.g., less than 50 students per group) to assess the influence of sample size on the results.

Risk of bias:

- Blinding: An analysis will be performed excluding studies where participants and / or researchers were not blinded to the intervention group (single-blind or open-label studies). Blinding can reduce bias due to expectations.
- Allocation concealment: Studies where the method of assigning participants to groups (intervention vs. control) was not adequately concealed will be excluded. Proper concealment helps to prevent selection bias.

- Funding source: An analysis will be performed excluding studies funded by entities with a vested interest in the EdTech intervention being evaluated. This can help to mitigate potential bias stemming from funding source.

Study characteristics:

- Publication status: An analysis will be conducted excluding unpublished studies to assess potential publication bias. Unpublished studies with negative findings might be less likely to be reported.
- Student population: Subgroup analyses will be conducted by student characteristics (e.g., grade level, socioeconomic status) to explore potential variation in the effectiveness of EdTech interventions across different populations.

6. Reporting

6.1. Quality of the evidence base

To establish the quality of the evidence base following the analytical and synthesis stages, we will employ the use of the GRADE Evidence to Decision frameworks (Grading of Recommendations Assessment, Development and Evaluation), which is a widely adopted tool used to assess the quality of evidence for making research-informed recommendations. It assesses the quality of a body of evidence based on outcomes, rather than papers.

We will use the established four levels of certainty ratings, where:

> **Very low** = The true effect is probably markedly different from the estimated effect
>
> **Low** = The true effect might be markedly different from the estimated effect
>
> **Moderate** = The authors believe that the true effect is probably close to the estimated effect
>
> **High** = The authors have a lot of confidence that the true effect is similar to the estimated effect.

The GRADE domains for rating are:

1. Risk of bias
2. Imprecision
3. Inconsistency
4. Indirectness
5. Publication bias

As this is a subjective process, this will be double-coded, where two members of the research team will code the bodies of evidence independently and discuss discrepancies in a meeting. If they cannot come to an agreement, a third (senior) reviewer will join the discussion and make a decision.

An overall GRADE rating can be applied to each body of evidence across the identified outcomes by taking the lowest quality of evidence from all outcomes that are crucial to decision-making and recommendations. This process will allow us to make confident recommendations based on the strength of the evidence base.

6.2. Published outputs

As an organisation, we are committed to promoting the publication, dissemination, and use of our research across a range of platforms and networks.

For this project, we will publish the research protocols and the bespoke coding tool on the Open Science Framework, the EEF website, and Open Development & Education's public-facing Evidence Library in May 2024. Then, we will present emerging findings from the systematic review and the structured community review to the EEF at the end of November 2024.

Finally, we will submit a peer-reviewed report for publication on the EEF's website and Open Development & Education's public-facing Evidence Library in February 2025.

The EEF may use the findings and the report to support the design and development of school-facing outputs, make updates to the Teaching and Learning Toolkit, and inform their future research and grant-making plans.

7. Methodological risks

The methodological risks associated with the three-stage approach outlined in this protocol are summarised in Table 6, along with our plans to mitigate these.

Table 6. Methodological risks

Risk description	**Probability (L/M/H)**	**Impact (L/ M/H)**	**Mitigation actions**
Selection bias	L	H	Clearly define inclusion and exclusion criteria in advance and follow them rigorously. Use a systematic search strategy to minimise selection bias. Engaging peers and peer review to ensure impartiality.
Publication bias	L	H	Employ methods such as Egger's test / Knapp-Hartung / Hartung-Knapp-Sidik-Jonkman method to assess and address publication bias
Heterogeneity	M	H	Use appropriate statistical techniques (e.g., random-effects models) to account for heterogeneity. Potentially conduct subgroup analyses or meta-regression to explore sources of heterogeneity (dependent on findings from heterogeneity tests — to be outlined in a future addendum that will be published before analysis).
Quality of studies	H	H	Assess the quality of included studies using established tools. Consider sensitivity analyses that exclude lower-quality studies.
Data availability	L	H	Contact authors for missing data whenever possible. Use sensitivity analyses.
Bias in reporting	M	H	Check for selective outcome reporting and, if necessary, contact authors for additional data or information.

Risk description	Probability (L/M/H)	Impact (L/ M/H)	Mitigation actions
Conflict of Interest	L	M	Disclose any conflicts of interest in the meta-analysis report. Consider sensitivity analyses that exclude studies with potential conflicts.
Overgeneralisation	L	H	Clearly define the scope and limitations of the meta-analysis in the discussion section of the report.

8. Timeline

Table 7. Timeline of project activities

Date	Activity	Staff responsible
April 2024	Submission of final research protocols	Hannah Walker and Chris McBurnie
April 2024	Submission of bespoke coding tool	Hannah Walker and Chris McBurnie
August 2024	Submission of background data merged with EEF database	Bethany Huntington and Christopher Klune
October 2024	Completion of systematic literature review with meta-analysis	Bethany Huntington and Christopher Klune
November 2024	Completion of structured community review	Bethany Huntington and Christopher Klune
December 2024	Submission of first draft of report	Bethany Huntington and Christopher Klune
February 2025	Submission of final draft of report	Bethany Huntington and Christopher Klune

9. Team

The research team will consist of the following individuals, with their exact roles and outputs detailed in Table 8.

Table 8. Research team

Team member	Role
Dr Björn Haßler (Open Development & Education)	Björn will work as the **Principal Investigator,** with responsibility for managing and overseeing the study. **Associated Outputs:** final research protocols; bespoke coding tool; background data; systematic literature review with meta-analysis; structured community review; first and final draft of reports.
Hannah Walker (Open Development & Education)	Hannah will act as **Co-Investigator**, working with Björn to oversee the design and delivery of the project. **Associated Outputs:** final research protocols; bespoke coding tool; background data; systematic literature review with meta-analysis; structured community review; first and final draft of reports.
Chris McBurnie (Open Development & Education)	During the inception phase, Chris will act as the project **Research Lead (Interim)** and focus on developing the conceptual framework, methodological approach, research protocols, and bespoke codebook. **Associated Outputs:** final research protocols; bespoke coding tool
Bethany Huntington (Open Development & Education)	From April, Bethany will work as the project **Research Lead**. In this role, Bethany will lead the systematic review and meta-analysis and the structured community review. **Associated Outputs:** background data; systematic literature review with meta-analysis; structured community review; first and final draft of reports. *Concluding PhD – available from April 2024.*

Team member	Role
Christopher Klune (Open Development & Education)	From April, Christopher will work as the project **Research Co-Lead**. In this role, Christopher will co-lead the systematic review and meta-analysis and the structured community review. **Associated Outputs:** background data; systematic literature review with meta-analysis; structured community review; first and final draft of reports.
Hassan Mansour (Open Development & Education)	Hassan will work as the **Technology Specialist**, overseeing the development and use of artificial intelligence tools for automated searches. **Associated Outputs:** bespoke coding tool; systematic literature review with meta-analysis.
Gemma Bennett (Open Development & Education)	Gemma represents the profession of educational psychology in this work, ensuring linkages with practising educational psychologists, particularly in deprived areas. *Concluding PhD – available from January 2024.*
Dr Louis Major (University of Manchester)	Louis will serve as a **Specialist Advisor (Programme Design)**, supporting the preparation of the protocols with a focus on the systematic literature review and structured community review. **Associated Outputs:** final research protocols; systematic literature review with meta-analysis; structured community review; first and final draft of reports.
Dr Aditi Bhutoria (Indian Institute of Management, Calcutta)	Aditi will serve as a **Specialist Advisor (Quantitative Methods)**, supporting the team to design and execute the meta-analysis. **Associated Outputs:** systematic literature review with meta-analysis; first and final draft of reports.
Evette Ferrao (Open Development & Education)	Evette will work as **Project Administrator,** supporting overall project management (e.g., invoicing, work plans, meetings).

10. Conflicts of interest

Researchers at Open Development & Education will conduct this EEF-funded study. The views expressed in this document do not necessarily reflect the views of EEF.

All authors declare no conflicts of interest.

11. References

This bibliography is available digitally in our evidence library at https://docs.opendeved.net/lib/2I2GT22T

Abrami, P. C., Lysenko, L., & Borokhovski, E. (2020). The effects of ABRACADABRA on reading outcomes: An updated meta-analysis and landscape review of applied field research. *Journal of Computer Assisted Learning*, *36*(3), 260–279. https://doi.org/10.1111/jcal.12417. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12417. (details)

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Review of Educational Research*, *82*(4), 436–476. https://doi.org/10.3102/0034654312458162. Available from http://journals.sagepub.com/doi/10.3102/0034654312458162. (details)

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, *126*, 334–345. https://doi.org/10.1016/j.compedu.2018.07.021. Available from https://linkinghub.elsevier.com/retrieve/pii/S0360131518302045. (details)

Alison, C. (2023). *Guidance note on using implementation research in education*. https://www.edu-links.org/sites/default/files/media/file/Guidance_Note_on_Using_Implementation_Research_in_Education.pdf. (details)

Archer, K., Savage, R., Sanghera-Sidhu, S., Wood, E., Gottardo, A., & Chen, V. (2014). Examining the effectiveness of technology use in classrooms: A tertiary meta-analysis. *Computers & Education*, *78*, 140–149. https://doi.org/10.1016/j.compedu.2014.06.001. Available from https://www.sciencedirect.com/science/article/pii/S036013151400133X. (details)

BaiDoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.61969/jai.1337500. Available from https://dergipark.org.tr/en/doi/10.61969/jai.1337500. (details)

Barzelay, M. (2007). Learning from Second-Hand Experience: Methodology for Extrapolation-Oriented Case Research. *Governance*, *20*(3), 521–543. https://doi.org/10.1111/j.1468-0491.2007.00369.x. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0491.2007.00369.x. (details)

Bates, M. A., & Glennerster, R. (2017). The Generalizability Puzzle. *Stanford Social Innovation Review*. https://ssir.org/articles/entry/the_generalizability_puzzle. (details)

Belland, B. R. (2017). Instructional Scaffolding: Foundations and Evolving Definition. In B. R. Belland (Ed.), *Instructional Scaffolding in STEM Education: Strategies and Efficacy Evidence* (pp. 17–53). Springer International Publishing. https://doi.org/10.1007/978-3-319-02565-0_2. (details)

Belland, B. R., Walker, A. E., & Kim, N. J. (2017). A Bayesian Network Meta-Analysis to Synthesize the Influence of Contexts of Scaffolding Use on Cognitive Outcomes in STEM Education. *Review of Educational Research*, *87*(6), 1042–1081. https://doi.org/10.3102/0034654317723009. (details)

Blei, D. M., & Lafferty, J. D. (2009a). *Visualizing Topics with Multi-Word Expressions* (No. arXiv:0907.1013). arXiv. http://arxiv.org/abs/0907.1013. (details)

Blei, D. M., & Lafferty, J. D. (2009b). Topic models. In *Text mining* (pp. 101–124). Chapman and Hall/CRC. https://www.taylorfrancis.com/chapters/edit/10.1201/9781420059458-12/topic-models-david-blei-john-la%EF%AC%80erty. (details)

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, *168*, 1–20. https://doi.org/10.1016/j.jpubeco.2018.08.007. Available from https://www.sciencedirect.com/science/article/pii/S0047272718301518. (details)

Bond, M. (2020). Schools and Emergency Remote Education during the COVID-19 Pandemic: A Living Rapid Systematic Review. *Asian Journal of Distance Education*, *15*(2), 191–247. https://eric.ed.gov/?id=EJ1285336. Available from https://eric.ed.gov/?id=EJ1285336. (details)

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: a systematic evidence map. *International Journal of Educational Technology in Higher Education*, *17*(1), 2. https://doi.org/10.1186/s41239-019-0176-8. (details)

Cabı, E. (2018). The Impact of the Flipped Classroom Model on Students' Academic Achievement. *The International Review of Research in Open and Distributed Learning*, *19*(3). https://doi.org/10.19173/irrodl.v19i3.3482. Available from http://www.irrodl.org/index.php/irrodl/article/view/3482. (details)

Castro, F. G., Barrera, M., & Holleran Steiker, L. K. (2010). Issues and Challenges in the Design of Culturally Adapted Evidence-Based Interventions. *Annual Review of Clinical Psychology*, *6*, 213–239. https://doi.org/10.1146/annurev-clinpsy-033109-132032. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4262835/. (details)

Chen, M.-H., Tseng, W.-T., & Hsiao, T.-Y. (2018). The Effectiveness of Digital Game-Based Vocabulary Learning: A Framework-Based View of Meta-Analysis. *British Journal of Educational Technology*, *49*(1), 69–77. https://doi.org/10.1111/bjet.12526. (details)

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001. Available from https://www.sciencedirect.com/science/article/pii/S1747938X13000031. (details)

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065. (details)

CooperGibson Research. (2022). *Implementation of education technology in schools and colleges*. Department for Education. (details)

D'Angelo, C., Rutstein, D., Harris, C., Haertel, G., Bernard, R., & Borokhovski, E. (2014). *Simulations for STEM Learning: Systematic Review and Meta-Analysis*. SRI Education. https://www.sri.com/wp-content/uploads/2021/12/simulations-for-stem-learning-brief.pdf. (details)

EPPI Centre. (2003). *Core Keywording Strategy*. https://eppi.ioe.ac.uk/CMS/Portals/0/PDF%20reviews%20and%20summaries/EPPI_Keyword_strategy.pdf. (details)

Education Endowment Foundation, & Durham University. (2022). *EEF Evidence Database Coding Guide – Main Data Extraction* (No. Version 3). https://d2tic4wvo1iusb.cloudfront.net/production/documents/toolkit/MDE_CodingGuide_V3_March2022-1.pdf. (details)

Evans, D. K., & Popova, A. (2016). What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. *The World Bank Research Observer*, *31*(2), 242–270. https://doi.org/10.1093/wbro/lkw004. (details)

Gerard, L., Matuk, C., McElhaney, K., & Linn, M. C. (2015). Automated, adaptive guidance for K–12 education. *Educational Research Review*, *15*, 41–58. https://doi.org/10.1016/j.edurev.2015.04.001. Available from https://www.sciencedirect.com/science/article/pii/S1747938X15000238. (details)

Glewwe, P., & Muralidharan, K. (2016). Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 5, pp. 653–743). Elsevier. https://doi.org/10.1016/B978-0-444-63459-7.00010-5.

Available from https://www.sciencedirect.com/science/article/pii/B9780444634597000105. (details)

Haßler, B. (2021m). *A mixed-methods evidence assessment for emergent fields of knowledge* (Working Paper No. 24). EdTech Hub. https://doi.org/10.53832/edtechhub.0003. Available from https://docs.opendeved.net/lib/E3W7NXR9. Available under Creative Commons Attribution 4.0 International (https://creativecommons.org/licenses/by/4.0/). (details)

Haßler, B., Adam, T., Allier-Gagneur, Z., Blower, T., Brugha, M., Damani, K., Hennessy, S., Hollow, D., Jordan, K., Martin, K., Megha-Bongnkar, G., Murphy, M., & Walker, H. (2021k). *Methodology for literature reviews* (Working Paper No. 10). EdTech Hub. https://doi.org/10.5281/zenodo.4557508. Available from https://docs.edtechhub.org/lib/2CKWI7RR. Available under Creative Commons Attribution 4.0 International. (details)

Haßler, B., El-Serafy, Y., Khalayleh, A., & McBurnie, C. (2021). *Systemic Mixed-Methods Research — a conceptual framework for scalable EdTech research*. EdTech Hub. https://doi.org/10.53832/edtechhub.0001. Available from https://docs.edtechhub.org/lib/TD2ASC4G. (details)

Haßler, B., Haseloff, G., Adam, T., Akoojee, S., Allier-Gagneur, Z., Ayika, S., Bahloul, K., Kigwilu, P. C., Costa, D. D., Damani, K., Gordon, R., Idris, A., Iseje, F., Jjuuko, R., Kagambèga, A., Khalayleh, A., Konayuma, G., Kunwufine, D., Langat, K., … Winkler, E. (2020a). *Technical and Vocational Education and Training in Sub-Saharan Africa: A Systematic Review of the Research Landscape* (1st ed.). VET Repository, Bundesinstitut für Berufsbildung, Bonn, Germany. https://doi.org/10.5281/zenodo.4288831. Available from https://docs.opendeved.net/lib/ZEDIHF57. Available under Creative Commons Attribution (CC BY 4.0). (details)

Haßler, B., McBurnie, C., Walker, H., Klune, C., & Huntington, B. (2024). *Keyword Inventory for a systematic literature review with meta-analysis: Understanding Quality Characteristics of EdTech Interventions and Implementation for Disadvantaged Pupils* [Other]. Open Development & Education. https://doi.org/10.53832/opendeved.1075. Available from https://docs.opendeved.net/lib/RNQSKET7. Available under Creative Commons Attribution 4.0 International. (details)

Haßler, B., McBurnie, C., Walker, H., Klune, C., & Huntington, B. (2024). *Supplementary coding and data extraction tool for a systematic review with meta-analysis: Understanding quality characteristics of EdTech interventions and implementation for disadvantaged pupils* (Understanding Quality Characteristics of EdTech Interventions and Implementation for Disadvantaged Pupils No. 2). Open Development & Education. https://doi.org/10.53832/opendeved.1115. Available from https://docs.opendeved.net/lib/G5P9ZNRI. Available under Creative Commons Attribution 4.0 International. (details)

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials – the gold standard for effectiveness research. *International Journal of Obstetrics & Gynaecology*, *125*(13), 1716–1716. https://doi.org/10.1111/1471-0528.15199. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/1471-0528.15199. (details)

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing Fidelity with Flexibility and Fit: What Do We Really Know about Fidelity of Implementation in Schools? *Exceptional Children*, *79*(3), 181–193. https://doi.org/10.1177/001440291307900204. (details)

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation. Full Report*. Education Endowment Foundation. https://eric.ed.gov/?id=ED612174. Available from https://eric.ed.gov/?id=ED612174. (details)

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/10.1111/ejed.12533. Available from https://onlinelibrary.wiley.com/doi/10.1111/ejed.12533. (details)

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F. K., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/efi-180221. (details)

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119–135. https://doi.org/10.1007/s13194-011-0038-2. Available from http://link.springer.com/10.1007/s13194-011-0038-2. (details)

Karboul, A. (2020, December 4). COVID-19 put 1.6 billion children out of school. Here's how to upgrade education post-pandemic. *World Economic Forum*. https://www.weforum.org/agenda/2020/12/covid19-education-innovation-outcomes/. (details)

Kerwin, J. T., & Thornton, R. L. (2021). Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures. *The Review of Economics and Statistics*, *103*(2), 251–264. https://doi.org/10.1162/rest_a_00911. (details)

Klune, C., Huntington, B., Touiyate, A., Kay, J., Zaman, M., Mansour, H., & Haßer, B. (2024). *Scaling AI-driven evidence synthesis for targeting new education research: Accelerating progress towards SDG4* [Extended abstract submitted to the What Works Hub for

Global Education Conference 2024]. Open Development & Education, Education Endowment Foundation. https://doi.org/10.53832/opendeved.1132. Available from https://docs.opendeved.net/lib/P527M7BW. Available under Creative Commons Attribution 4.0 International. (details)

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420. (details)

Kunkel, A. (2015). *The Effects of Computer-Assisted Instruction in Reading: A Meta-Analysis* [University of Minnesota]. http://conservancy.umn.edu/handle/11299/175221. (details)

Leviton, L. C. (2017). Generalizing about Public Health Interventions: A Mixed-Methods Approach to External Validity. *Annual Review of Public Health*, *38*(1), 371–391. https://doi.org/10.1146/annurev-publhealth-031816-044509. (details)

Lewin, C., Smith, A., Morris, S., & Craig, E. (2019). *Using Digital Technology to Improve Learning: Evidence Review*. Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/production/documents/guidance/Using_Digital_Technology_to_Improve_learning_Evidence_Review.pdf?v=1698320282. Available from https://d2tic4wvo1iusb.cloudfront.net/production/documents/guidance/Using_Digital_Technology_to_Improve_learning_Evidence_Review.pdf?v=1698320282. (details)

Lin, H. (2014). Establishing an Empirical Link between Computer-Mediated Communication (CMC) and SLA: A Meta-Analysis of the Research. *Language Learning & Technology*, *18*(3), 120–147. (details)

Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism Experiments and Policy Evaluations. *Journal of Economic Perspectives*, *25*(3), 17–38. https://doi.org/10.1257/jep.25.3.17. Available from https://www.aeaweb.org/articles?id=10.1257/jep.25.3.17. (details)

Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology*, *52*(5), 1935–1964. https://doi.org/10.1111/bjet.13116. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13116. (details)

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania*. *The Quarterly Journal of Economics*, *134*(3), 1627–1673. https://doi.org/10.1093/qje/qjz010. (details)

McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2016). *ABRA: Online Reading Support. Evaluation Report and Executive Summary*. Education Endowment Foundation. https://eric.ed.gov/?id=ED581099. Available from https://eric.ed.gov/?id=ED581099. (details)

McNeish, D. (2017). Challenging Conventional Wisdom for Multivariate Statistical Models With Small Samples. *Review of Educational Research*, *87*(6), 1117–1151. https://doi.org/10.3102/0034654317727727. Available from http://journals.sagepub.com/doi/10.3102/0034654317727727. (details)

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature. *Teachers College Record: The Voice of Scholarship in Education*, *115*(3), 1–47. https://doi.org/10.1177/016146811311500307. Available from http://journals.sagepub.com/doi/10.1177/016146811311500307. (details)

Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van Den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, *20*(6), 559–572. https://doi.org/10.1080/13645579.2016.1252189. Available from https://www.tandfonline.com/doi/full/10.1080/13645579.2016.1252189. (details)

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, *350*, h1258. https://doi.org/10.1136/bmj.h1258. Available from https://www.bmj.com/content/350/bmj.h1258. Available under © Moore et al 2015. This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/.. (details)

ONS. (2023). *A level and other 16 to 18 results, Academic year 2022/23*. https://explore-education-statistics.service.gov.uk/find-statistics/a-level-and-other-16-to-18-results/2022-23#dataBlock-5c6d58a3-56e3-4ffe-6c95-08dbd1502a35-tables. (details)

ONS. (2023). *Key stage 1 and phonics screening check attainment, Academic year 2022/23*. https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-1-and-phonics-screening-check-attainment/2022-23. (details)

ONS. (2023). *Key stage 2 attainment, Academic year 2022/23*. https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment/2022-23. (details)

ONS. (2023). *Key stage 4 performance, Academic year 2022/23*. https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-4-performance-revised. (details)

Outhwaite, L. A., Early, E., Herodotou, C., & Van Herwegen, J. (2023). Understanding how educational maths apps can enhance learning: A content analysis and qualitative comparative analysis. *British Journal of Educational Technology*, *54*(5), 1292–1313. https://doi.org/10.1111/bjet.13339. Available from https://bera-journals.onlinelibrary.wiley.com/doi/10.1111/bjet.13339. (details)

Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2019). Raising early achievement in math with interactive apps: A randomized control trial. *Journal of Educational Psychology*, *111*(2), 284. https://doi.org/10.1037/edu0000286. (details)

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017). Closing the gap: efficacy of a tablet intervention to support the development of early mathematical skills in UK primary school children. *Computers & Education*, *108*, 43–58. https://doi.org/10.1016/j.compedu.2017.01.011. (details)

Outhwaite, L., Ang, L., Herbert, E., Sumner, E., & Van Herwegen, J. (2023). *Technology and Learning for Early Childhood and Primary Education* [Report]. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386108.locale=en. (details)

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, *10 Suppl 1*, 21–34. https://doi.org/10.1258/1355819054308530. (details)

Piper, B., Jepkemei, E., Kwayumba, D., & Kibukho, K. (2015). Kenya's ICT Policy in Practice: The Effectiveness of Tablets and E-readers in Improving Student Outcomes. *FIRE: Forum for International Research in Education*, *2*(1). https://doi.org/10.18275/fire201502011025. Available from https://preserve.lehigh.edu/fire/vol2/iss1/2. (details)

Pokhrel, S., & Chhetri, R. (2021). A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. *Higher Education for the Future*, *8*(1), 133–141. https://doi.org/10.1177/2347631120983481. Available from http://journals.sagepub.com/doi/10.1177/2347631120983481. (details)

Pritchett, L., & Sandefur, J. (2015). Learning from Experiments When Context Matters. *American Economic Review*, *105*(5), 471–475. https://doi.org/10.1257/aer.p20151016. Available from https://www.aeaweb.org/articles?id=10.1257/aer.p20151016. (details)

Pustejovsky, J. E., & Tipton, E. (2021). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1007/s11121-021-01246-3. (details)

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-Analysis With Complex Research Designs: Dealing With Dependence From Multiple Measures and Multiple Group Comparisons. *Review of Educational Research*, *84*(3), 328–364. https://doi.org/10.3102/0034654313500826. Available from http://journals.sagepub.com/doi/10.3102/0034654313500826. (details)

Sharpe, D., & Poets, S. (2020). Meta-analysis as a response to the replication crisis. *Canadian Psychology/Psychologie Canadienne*, *61*(4), 377. https://doi.org/10.1037/cap0000215. Available from https://psycnet.apa.org/record/2020-58586-001. (details)

Sharples, J., Albers, B., Fraser, S., & Kime, S. (2019). *Putting Evidence to Work: A School's Guide to Implementation*. https://d2tic4wvo1iusb.cloudfront.net/eef-guidance-reports/implementation/EEF-Active-Ingredients-Summary.pdf?v=1635355218. Available from https://d2tic4wvo1iusb.cloudfront.net/eef-guidance-reports/implementation/EEF-Active-Ingredients-Summary.pdf?v=1635355218. (details)

Sims, S., & Fletcher-Wood, H. (2018). *Characteristics of effective teacher professional development: what we know, what we don't, how we can find out*. (details)

Sims, S., Fletcher-Wood, H., O'Mara, A., Cottingham, S., Stansfield, C., Herwegen, J. V., & Anders, J. (2021). *What are the Characteristics of Effective Teacher Professional Development? A Systematic Review and Meta-analysis*. (details)

Sokolowski, A., Li, Y., & Willson, V. (2015). The effects of using exploratory computerized environments in grades 1 to 8 mathematics: a meta-analysis of research. *International Journal of STEM Education*, *2*(1), 8. https://doi.org/10.1186/s40594-015-0022-z. (details)

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, *105*(4), 970–987. https://doi.org/10.1037/a0032447. (details)

Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, *94*, 252–275. https://doi.org/10.1016/j.compedu.2015.11.008. Available from https://www.sciencedirect.com/science/article/pii/S0360131515300804. (details)

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and Pitfalls of Multimedia and Interactive Features in Technology-Enhanced Storybooks: A Meta-Analysis. *Review of Educational Research*, *85*(4), 698–739. https://doi.org/10.3102/0034654314566989. (details)

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and Pitfalls of Multimedia and Interactive Features in Technology-Enhanced Storybooks: A Meta-Analysis. *Review of Educational Research*, *85*(4), 698–739. https://doi.org/10.3102/0034654314566989. (details)

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019a). A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods*, *10*(2), 161–179. https://doi.org/10.1002/jrsm.1338. Available from https://onlinelibrary.wiley.com/doi/10.1002/jrsm.1338. (details)

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019b). Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods*, *10*(2), 180–194. https://doi.org/10.1002/jrsm.1339. Available from https://onlinelibrary.wiley.com/doi/10.1002/jrsm.1339. (details)

UK Government. (2019). *Realising the potential of technology in education: A strategy for education providers and the technology industry*. https://assets.publishing.service.gov.uk/media/5ca360bee5274a77d479facc/DfE-Education_Technology_Strategy.pdf. (details)

UK Government. (2023). *Pupil premium: overview*. GOV.UK. https://www.gov.uk/government/publications/pupil-premium/pupil-premium. (details)

UNCTAD. (2023). *Technology and Innovation Report*. https://unctad.org/system/files/official-document/tir2023_en.pdf. (details)

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, *47*, 1274–1294. https://doi.org/10.3758/s13428-014-0527-2. Available from https://link.springer.com/article/10.3758/s13428-014-0527-2. (details)

Verbruggen, S., Depaepe, F., & Torbeyns, J. (2021). Effectiveness of educational technology in early mathematics education: A systematic literature review. *International Journal of Child-Computer Interaction*, *27*, 100220. https://doi.org/10.1016/j.ijcci.2020.100220. Available from https://www.sciencedirect.com/science/article/pii/S2212868920300386. (details)

Vicentini, L., Day, L., Gill, V., Lillis, J., Komers, S., & Olausson, N. (2022). *Future opportunities for education technology in England*. Department for Education. (details)

Vivalt, E. (2020). How Much Can We Generalize From Impact Evaluations? *Journal of the European Economic Association*, *18*(6), 3045–3089. https://doi.org/10.1093/jeea/jvaa019. (details)

Walker, H., Moustafa, N., & Haßler, B. (2023). *EdTech Impact in Education – An Analysis Prepared for the Education Endowment Foundation*. OpenDevEd. https://doi.org/10.53832/opendeved.0294. Available from https://docs.opendeved.net/lib/4PD4PMAF. Available under Creative Commons Attribution 4.0. (details)

West, M. (2023). *An ed-tech tragedy? Educational technologies and school closures in the time of COVID-19* (K. Davidson & R. Yaghmour, Eds.). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386701. (details)

Williams, M. J. (2020). External Validity and Policy Adaptation: From Impact Evaluation to Policy Design. *The World Bank Research Observer*, *35*(2), 158–191. https://doi.org/10.1093/wbro/lky010. Available from https://academic.oup.com/wbro/article/35/2/158/5542577. (details)

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311. Available from http://doi.apa.org/getdoi.cfm?doi=10.1037/a0031311. (details)

Xiao, Y., & Watson, M. (2017). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, *39*, 0739456X1772397. https://doi.org/10.1177/0739456x17723971. Available from https://doi.org/10.1177/0739456X17723971. (details)

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, *2021*, 1–18. https://doi.org/10.1155/2021/8812542. Available from https://www.hindawi.com/journals/complexity/2021/8812542/. (details)

Zheng, B., Warschauer, M., Lin, C.-H., & Chang, C. (2016). Learning in One-to-One Laptop Environments: A Meta-Analysis and Research Synthesis. *Review of Educational Research*, *86*(4), 1052–1084. https://doi.org/10.3102/0034654316628645. (details)

Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: a meta-analysis. *Asia Pacific Education Review*, *17*(2), 187–202. https://doi.org/10.1007/s12564-016-9426-9. (details)

Zheng, L., Li, X., Tian, L., & Cui, P. (2018). The effectiveness of integrating mobile devices with inquiry-based learning on students’ learning achievements: a meta-analysis.

International Journal of Mobile Learning and Organisation, *12*(1), 77–95.
https://doi.org/10.1504/IJMLO.2018.089238. Available from
https://www.inderscienceonline.com/doi/abs/10.1504/IJMLO.2018.089238. (details)

12. Annexes

12.1. Annex 1. Teacher Interview Schedule

Phase 1: Stakeholder Engagement Teacher Interview Questions

Scope and Purpose

Technology is playing an increasingly significant role in the daily lives of teachers and learners. In England, for example, over two-thirds of schools have introduced, upgraded, or increased their use of technology since 2022 (↑CooperGibson Research, 2022). In this context, many teachers are expected to select and implement new technologies without guidance on what works for what students.

We are working with the Education Endowment Foundation to study how the use of educational technology can raise student attainment across subjects, with a focus on disadvantaged pupils. This study will consist of three phases:

1. Stakeholder engagement
2. Literature review
3. Community review of the gathered evidence

This survey is part of the stakeholder engagement. Data gathered from the survey will inform our conceptual and methodological approach to the rest of the project. Data will also be used as part of the research for a masters thesis.

We truly appreciate your participation in the survey. Your insights and experiences are deeply valuable, as they will help us refine our understanding of key concepts in this project and improve our approach to the literature review. If you have any questions about the interview, please let us know. We can begin when you are ready.

Demographics		
Q No.	**Question**	**Probes/comments**
1	How many years of teaching experience do you have?	
2	We'd like to better understand your teaching role: ● Primary/secondary ● What subjects do you currently teach? ● Do you have any other additional roles at your school?	**Primary school** Ks 1, ks2 — which year do you teach? **Secondary school — years 7-11. Age 11-16 or 18.** Ks3 (years 7-9) 11-15 years old K34 (years 10-11 GCSE year) 14-16 years old KS5 (Years 12-13 A Levels) 16-18 years old.
Interview Questions		
Experiences with Educational Technology and Learning: *Educational technology (EdTech) is an umbrella term, encompassing a range of technologies and a range of meanings for different people. For this study, we are adopting a broad definition of EdTech which I'd like to share with you: "Education technology (EdTech) refers to the practice of using technology to support teaching and the effective day-to-day management of education institutions. It includes hardware (such as tablets, laptops or other digital devices), and digital resources, software and services that help aid teaching, meet specific needs, and help the daily running of education institutions (such as management information systems, information sharing platforms and communication tools)." You may have similar or different ideas when it comes to defining or understanding EdTech. With that in mind, we are mostly interested in your experiences with EdTech and student learning. So, the first question I'd like to ask is…*		

1	What types of EdTech do you use in your school or classroom to support student learning? ● Which ones do you find are most effective? Why do you use these particular tools or methods?	Comment: Ensure there is the focus on student learning. Probes: Software? hardware? apps?
2	Can you provide examples of when you have used EdTech in the classroom to effectively support student learning? Please use concrete examples (e.g. think about specific experiences). ● Why do you think it was effective (e.g. what factors were in place that made it effective?)	Probes: Why did you choose to use this tech? What evidence demonstrated it supported student learning? Can you describe the teaching strategies or approaches you used to integrate EdTech? How did students respond? How did this support student attainment?
3	Can you provide examples of when you have used EdTech in the classroom where it has not effectively supported student learning? Please use concrete examples (e.g. think about specific experiences). ● What were the main barriers to student learning here? Implementation	Probes : Can you describe the strategies or approaches you used to integrate EdTech? How did students respond? How did this support student attainment?

Factors and Disadvantaged Students: *There have been some really interesting insights from these questions. With these experiences in mind, I'd now like to ask…*		
5	What factors do you think influence the effective use of EdTech in improving student learning? ● Examples of these factors?	Comment: Open up the word 'factors' if needed — components, elements, considerations, observations, things from your experiences
6	With those in mind, I'd like to narrow the focus a bit. Are there any additional factors to consider when using EdTech with disadvantaged students? ● For example, additional barriers? Different considerations? Anything additional factors needed to be present or implemented? Give specific examples. ● As it relates to their learning	**Comment:** If needed, clarify disadvantaged as students that access PP or free school meals
Educational Technology and Teachers: *Thank you for those insights. Now, I'd like to shift to some questions focused on how you think EdTech has impacted your teaching, so…*		
7	How has EdTech, if at all, influenced or changed your teaching practice? ● Examples?	PROBES: Positively, negatively? How do these changes make you feel?
8	How do you personally assess the effectiveness of an EdTech tool?	

9	Where do you see future opportunities or challenges in using EdTech to support student learning?	

Educational technology

1. We would like to present to you a definition of educational technology and ask some questions related to it: "Education technology (EdTech) refers to the practice of using technology to support teaching and the effective day-to-day management of education institutions. It includes hardware (such as tablets, laptops or other digital devices), and digital resources, software and services that help aid teaching, meet specific needs, and help the daily running of education institutions (such as management information systems, information sharing platforms and communication tools)"
    a. What thoughts do you have about this definition? *Allow participant to answer, probe if necessary*
        i. Would you change anything about this definition? If so, what?
        ii. In your view, are there any aspects of this definition that you would critique? If so, please explain.
        iii. In your view, are there any aspects of this definition that you would expand upon? If so, please explain.

Mechanisms of change

1. We have conceptualised EdTech interventions at 4 levels: mechanisms, building blocks, models, and interventions. We'd like to get your thoughts on our definition of mechanisms of change, that is, the smallest component that underpins these levels of EdTech interventions. We define a mechanism of change as "A technological or non-technological feature (e.g., practice, behaviour, or activity) of an intervention that contributes to improved student attainment."
    a. What thoughts do you have about this definition? *Allow participant to answer, prompt with follow-ups if necessary*

iv. Would you change anything about this definition? If so, what?
v. In your view, are there any aspects of this definition that you would critique? If so, please explain.
vi. In your view, are there any aspects of this definition that you would expand upon? If so, please explain.

Categories and sub-categories for literature review

1. We will be doing a literature review on studies related to the impact of different uses of educational technology on student attainment. For this, we have made a list of categories and sub-categories to help us identify relevant studies. We would like to present you with these and ask a couple of questions to help us refine it. *Take the participant through the table. Depending on timing and flow, ideally go through each category one by one using these questions. If this is not possible, ask about the table generally.*
   a. Would you add or change anything about the categories?
   b. Would you add or change anything about the sub-categories?
   c. What keywords come to mind when you think of any of the categories or sub-categories?

Category	**Description**	**Sub-Categories**
Technology	The technology under examination in the study	Accessibility and Inclusion
		Communication and Social Media
		Device
		Electronic Resource
		Model of Delivery
		Software

		Technology
Disadvantage	The type or types of disadvantage under examination in the study	Free-school Meals
		Local Authority Care
		Special Educational Needs and Difficulties (SEND)
		Roma
		Traveller of Irish Heritage
		Black Caribbean
		Low-performing Local Authority
Outcomes and Impact		Accessibility
		Attainment
Research Methods	The research designs and methods of the study	Research Design
		Research Method

Geography	The country or region where the study was carried out	Area
		Country
Education Setting	The educational setting where the study was carried out	Distance Learning
		Higher Education
		Home
		Post-compulsory
		Distance Learning
		Higher Education
		Pre-Primary Education
		Primary and Secondary School
		Primary Education
		Pupil Referral Unit
		Secondary School
Instructional Domain	The focus population of the study	Generic
		Language
		Mathematics
		Other subjects

		Science

12.2. Annex 2. Teacher Survey Questions

Phase 1: Stakeholder Engagement Teacher Survey

Thank you for agreeing to complete this survey as part of our study on the characteristics of effective EdTech interventions for disadvantaged pupils. Below, you will find a brief overview of the scope and purpose of the project, as well as the questions.

We truly appreciate your participation in the survey. Your insights and experiences are deeply valuable, as they will help us refine our understanding of key concepts in this project and improve our approach to the literature review. If you have any questions about this survey, please contact: the Principal Investigator, Björn Haßler, on [email address], or the Co-Investigator, Hannah Walker, on [email address].

Scope and Purpose

Technology is playing an increasingly significant role in the daily lives of teachers and learners. In England, for example, over two-thirds of schools have introduced, upgraded, or increased their use of technology since 2022 (Cooper Gibson Research, 2022). In this context, many teachers are expected to select and implement new technologies without guidance on what works for what students.

We are working with the Education Endowment Foundation to study how the use of educational technology can raise student attainment across subjects, with a focus on disadvantaged pupils. This study will consist of three phases:

1. Stakeholder engagement
2. Literature review
3. Community review of the gathered evidence

This survey is part of the stakeholder engagement. Data gathered from the survey will inform our conceptual and methodological approach to the rest of the project.

Purpose

Survey

Demographics

1. How many years of teaching experience do you have?
    - 0-3 years
    - 4-6 years
    - 7-9 years
    - 10+ years
2. What year group(s) do you currently teach (select all that apply)?
    - Key Stage 1
    - Key Stage 2
    - Key Stage 3
    - Key Stage 4
3. What subject(s) do you currently teach (select all that apply)?
    - English
    - Mathematics
    - Sciences (including biology, chemistry, physics)
    - Humanities (including history, geography, philosophy, religions, performing arts, fine arts)
    - Other (please specify)
4. Do you have any other additional roles at your school?

Background on Educational Technology

1. Please list any forms of technology you use in your teaching to support student learning:
2. Describe how often you make use of educational technology in your teaching:
    a. Never (I have never used EdTech)
    b. Rarely (I usually use EdTech one or two times a month to support teaching)
    c. Occasionally (I usually use EdTech once per week to support teaching)
    d. Often (I usually use EdTech two or three times per week to support teaching)

    e. Always (I usually use EdTech daily or almost daily to support teaching)
3. Which best describes your teaching experience using EdTech?
    a. Overwhelmingly negative
    b. Mostly negative
    c. Neutral
    d. Mostly positive
    e. Overwhelmingly positive
4. Can you describe the specific steps or methods you take to effectively use educational technology to support student learning? You may describe an example from your own teaching.

Educational Technology and Learning

5. Can you briefly describe the most significant factors that you think influence the use of EdTech in improving student learning?

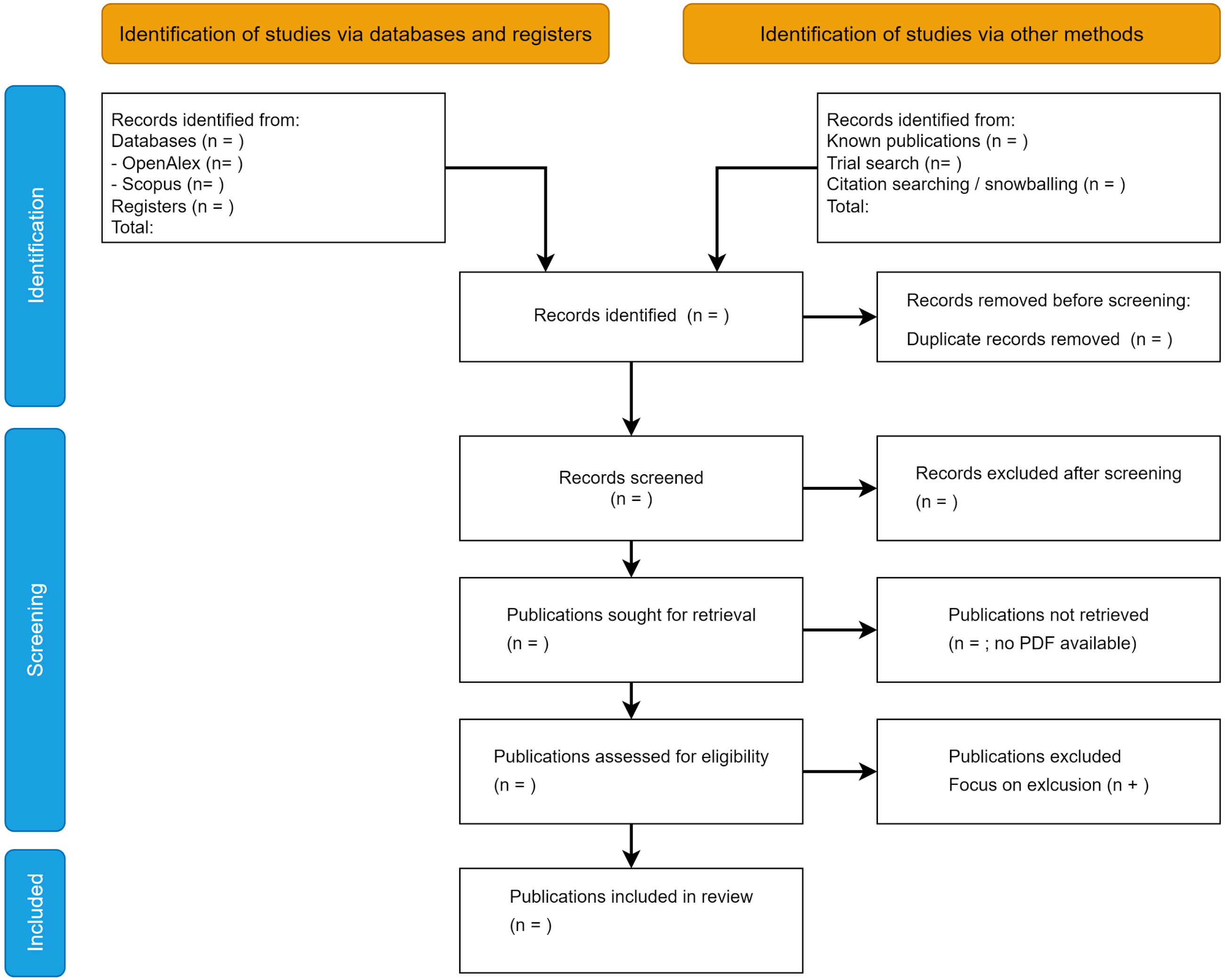
6. With these factors in mind, are there any additional considerations or differences when using EdTech with disadvantaged students? If so, can you describe these considerations and differences?

Educational Technology and Teaching

7. Can you briefly describe how EdTech, if at all, has influenced or changed your teaching practice?

8. (Optional) Is there anything else you would like to express about Edtech and improved student learning that you feel was not captured in the questions here?

12.3. Annex 3. PRISMA Diagram



12.4. Annex 4. Coding tool for abstract screening

- Date of Publication:
    - Was the paper published during or after 2011?
        - Yes
        - No
- Publication Type (Primary Focus):
    - Is the paper a peer-reviewed journal article, working paper, EEF evaluation report, or grey literature (from the list of relevant literature)?
        - Yes
        - No
- Research Design (Primary Focus):
    - Does the study utilise rigorous causal inference strategies (e.g., RCTs, quasi-experimental methods)?
        - Yes
        - No
- Research Design (Secondary Focus):
    - Does the study employ other rigorous research methods (e.g., in-depth qualitative studies, process evaluations)?
        - Yes
        - No
- Language of Publication:
    - Is the paper published in English?
        - Yes
        - No
- Geographic Focus:
    - Does the study focus on countries with 'high technological readiness'?
        - Yes
        - No
- Population (P):
    - Does the study focus on students in formal education in KS1–KS5?
    - Does the study focus on disadvantaged students in formal education in KS1–KS5 (according to the UK Government's definition)?
        - Yes
        - No
- Intervention (I):
    - Does the study examine interventions using any device or digital approach to support teaching and learning activities with the primary goal of improving student attainment?
        - Yes
        - No
- Comparison (C):
    - Does the study include a comparison group with no intervention, non-EdTech intervention, or waitlist intervention?
        - Yes
        - No

- Outcomes (O):
    - Does the study use tests to quantitatively measure student attainment in any curriculum subjects?
        - Yes
        - No

Based on the above coding, rank the paper as below:

- High (H): clearly satisfactory
- Medium (M): unclear or contentious
- Low (L): clearly unsatisfactory

12.5. Annex 5. Geographic focus and low-income schemes

For this study, we will focus on countries and territories that the United Nations Conference on Trade and Development classifies as exhibiting 'high' technological readiness (↑UNCTAD, 2023). These countries and territories are listed in the table below.

We are also interested in the measures of disadvantage specific to countries other than UK, that could be compared to the Free School Meals and Pupil Premium schemes, so have included this non-exhaustive list alongside the corresponding countries. This will help the research team identify relevant papers set in other countries during the screening process.

Countries	**Low-income schemes** *Name if specific policy, otherwise a description of who it applies to*
Australia	School Financial Assistance Scheme
Austria	
Belgium	Equal education opportunities policy — GOK (Gelijke Onderwijskansen)
Brazil	
Canada	
China	
Cyprus	Free school meals in primary schools for low-income households, children of asylum seekers, unaccompanied migrant children, and children under state guardianship.
Czechia	Funding scheme for free school meals for children aged 3-15 from lowest-income households.
Denmark	
Estonia	

Finland	
France	REP, REP+ L'Education Prioritaire Cantine Scolaire (School Canteen) Programme Provides free or subsidised meals based on family income
Germany	Lunch is reimbursed as part of the education and participation benefits of low-income households with children with basic income support for jobseekers, social assistance, asylum seekers benefits or supplementary child benefits or housing benefits.
Hong Kong	
Hungary	Iskolai Étkeztetési Program Free school meals in primary school and 50% reduction in secondary school for children receiving regular child protection benefits or in foster care.
Iceland	
Ireland	Delivering Equality of Opportunity in Schools DEIS
Israel	
Italy	
Japan	
Korea (Republic of)	
Latvia	

Luxembourg	
Malaysia	Program Pemberian Makanan Tambahan Supplementary feeding programme
Malta	
Netherlands	Educational Disadvantage Policy Onderwijs-Achterstanden Beleid OAB Gratis Schoolboeken Free school books programme to students from low-income families
New Zealand	
Norway	
Poland	
Portugal	
Russia	
Singapore	
Slovakia	
Slovenia	
Spain	
Sweden	
Switzerland	

United Arab Emirates	
United Kingdom	Pupil Premium Free School Meals
United States	Free and Reduced Price Lunch Program (FRPL)