TOWARDS A HOLISTIC UNDERSTANDING OF EVIDENCE: A WORKING PAPER

Natalia I. Kucirkova and David Dockterman

University of Stavanger and Harvard Graduate School of Education

International Centre for EdTech Impact, 2024

ISBN: 978-82-8439-288-2

Cite as: Kucirkova, N. & Dockterman, D. (2024). Towards a holistic understanding of evidence: a working paper. University of Stavanger. https://doi.org/10.31265/USPS.284

This work is licensed under a Creative Commons Attribution 4.0 International License.

Introduction

Research should spur responsible and effective innovation across domains, and K-12 EdTech is no exception. As two learning science researchers actively promoting evidence-based practices within the EdTech industry, Natalia (based in the UK and Norway) and David (based in the USA) have facilitated academia-industry partnerships at their respective universities. We have mentored young scholars interested in learning design and impact and trained education technology developers and entrepreneurs. We have focused on translating academic insights into market-aligned products grounded in learning sciences. We both have experience as independent academics evaluating products as well as engaging directly in commissioned consultancy work for commercial product development. In addition, David co-founded an early EdTech startup and worked in the publishing industry for decades.

Previous reports from the International Center for EdTech Impact focused on models of how academic research could serve EdTech developers/providers. In this paper, we take a broader perspective to highlight how members of the evidence ecosystem can align complementary goals and skill sets in service of measurable impact at scale.

Unfortunately, despite growing interest in research-industry connections, we recognize that deep, productive engagement between academics and EdTech providers remains uncommon. The two groups tend to operate independently of each other. Traditionally, researchers only connect to the EdTech industry by evaluating existing interventions or by providing generally available research or prototypes to bolster new product development. Additionally, academics conduct research that feeds into the development of ethical standards to address broad issues, such as algorithmic bias, social impact, privacy protection, and the risks of black box artificial intelligence, that at best, have only indirectly influenced EdTech product creation. Meanwhile, the EdTech industry has focused mainly on market research to identify and clarify what educators want, what the market can afford, and how customers and users like a product.

As long-time participants in the EdTech research and commercial worlds, we recognize the necessity for both academic and market types of research to produce interventions with a solid research and efficacy base that will also be used, and used effectively, in the field. This working paper takes a pragmatic approach and adopts a quality improvement perspective. We see an urgent need to integrate more research into EdTech development, and we observe a genuine industry interest in presenting evidence with integrity. However, we also note a lack of infrastructure to support that deeper integration of research and EdTech products.

We offer this working paper to all members of the K-12 EdTech evidence ecosystem - researchers, product developers, publishers, funders, policymakers, and, of course, educators. We seek to organize and spark shared dialogue about the meaning of “evidence of impact” in EdTech and how to incorporate evidence throughout the lifecycle of product discovery, development, and implementation at scale.

We suggest 3 paradigm shifts to drive the conversation: 1) Evidence of impact is multidimensional; 2) Evidence should feed an iterative impact process over time; 3) Rigor matters throughout the process. While members of the ecosystem have different incentives and metrics of success, they each bring different capabilities which, when aligned toward a shared goal of improved educational equity, can accelerate both the acquisition of sharable knowledge and wider and deeper impact.

The Current State: Divergent Views of EdTech Evidence

Following the global transition to online education during the COVID-19 pandemic, there has been a notable shift in narrative towards integrating evidence into an ecosystem that has historically undervalued, or at the very least, grappled with defining its impact on learners. Lots of money has been invested in educational technology (EdTech) historically, from magic lanterns to film projectors to televisions to computers (in multiple waves), each time failing to achieve the promise of revolutionary shifts in educational processes and outcomes. Currently, demands for evidence of impact are emerging from several parts of the K-12 ecosystem.

The U.S. federal government has set evidence standards in the Every Student Succeeds Act (ESSA). Social impact investors, including venture capitalists and philanthropies, have sought to capture the impact (along with the sustainability) of their investments. Philanthropic organisations, like The Jacobs Foundation, have promoted initiatives to encourage evidence collection for EdTech products. EdTech consumers - school administrators, teachers, parents, IT directors, and others involved in product selection and implementation in schools - have long considered the potential value of a program or intervention on student performance, engagement, and well-being, along with protection of privacy and sensitive information. Unfortunately, the varied perspectives of producers, consumers, and researchers are rarely aligned or fully realized.

One notable challenge in aligning the perspectives is a shared understanding around evidence and impact in EdTech. We need to broaden and clarify what we collectively mean by impact and how to measure it to bring coherence and rigor to the pursuit of demonstrable, positive change in the education of children, especially those who have been historically marginalized.

Evidence is used loosely in discussions around EdTech impact, often interchangeably with the words of effectiveness or efficacy. According to the Oxford Dictionary, "evidence" is "the facts, signs, or objects that make you believe that something is true". In EdTech, evidence is thus the "proof" that the use of a specific tool works (or will work), in a given context, for a given student or teacher.

There are various types of evidence, including reviews, research studies (published and conducted by independent researchers or internally by EdTech companies), feedback and testimonials from teachers or parents. These different types of evidence have different rigor or weight behind them, but they are all valid and important in determining whether something works.

Unsurprisingly, academics value and pursue a different kind of evidence than EdTech ventures do. The diagram below displays the typical evidence collection process from each perspective. Academics identify a fundable research question, carry out a study in a qualified set of contexts, collect and publish the results. The EdTech industry, on the other hand, typically conducts market research to identify a product opportunity that informs product development, marketing, and sales. Once the product is in use, evidence, often in the form of customer testimonials or correlational studies, is collected to support further sales. Occasionally the two worlds intersect, represented by the dotted lines in the diagram. Published studies may be used by industry to guide new products. Industry may commission research from academics to inform new products or validate products already in the market.

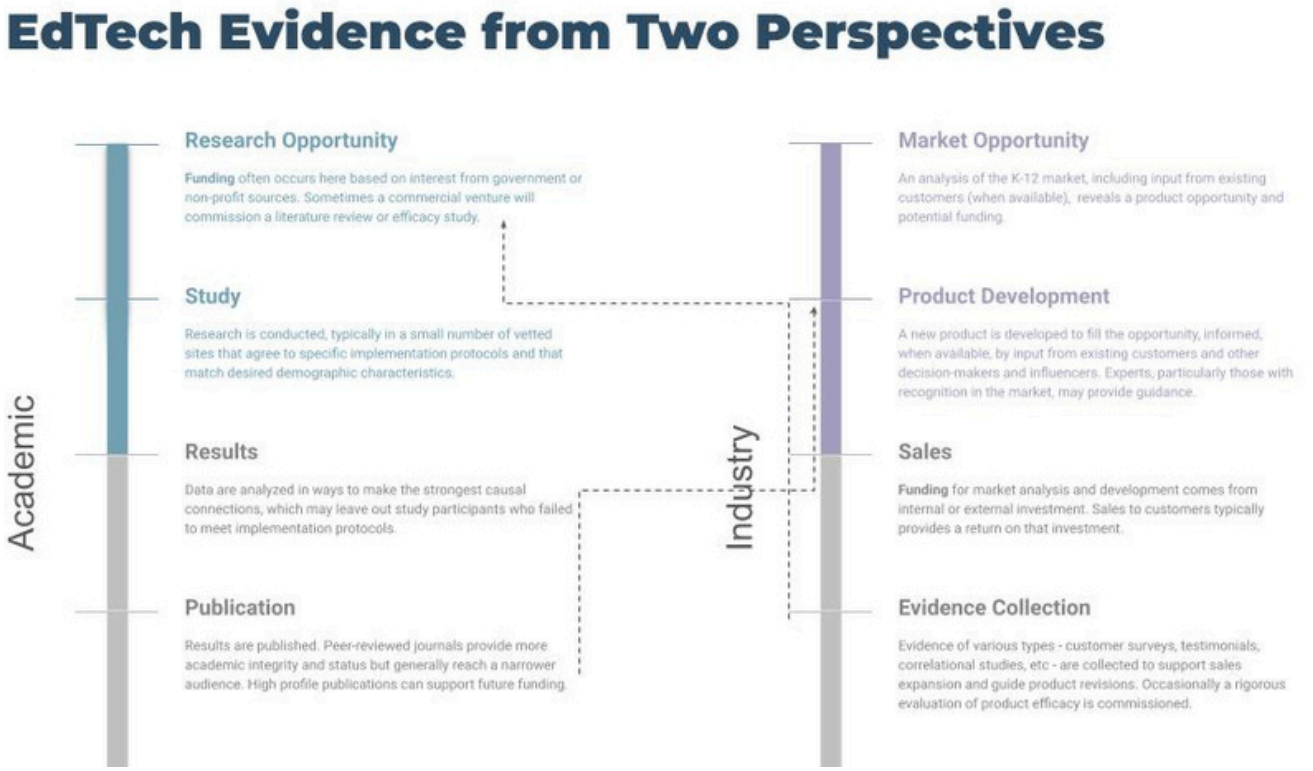

Figure 1: Academic and Industry Approach to Evidence (Summary of processes)

We readily acknowledge the generalizations in our representation. We have each been part of exceptions to the pathways depicted above, but those instances were decidedly exceptional. The professional incentives for researchers to publish papers and the sustainability pressures for EdTech ventures to acquire customers do not incentivize deep academic-industry collaboration. That lack of coordinated interaction is unfortunate, because evidence matters throughout the process of developing and implementing an intervention in K-12 settings. Evidence, for instance, should inform the design of EdTech products, boosting confidence that the intervention should work as intended across the target educational contexts. Evidence should also show that the product is working, providing insights into change and progress, while providing guidance on how to adapt the program to improve its effectiveness. Finally, evidence should be collected to show that the program contributed to promised outcomes, that it did work.

Each member of the evidence ecosystem can contribute to a more holistic, rigorous, and iterative process for creating and implementing effective educational programs. To align and activate the ecosystem for impact, we need some significant shifts in our thinking.

Three Paradigm Shifts

We propose three paradigmatic shifts in how impact is understood over time in EdTech:

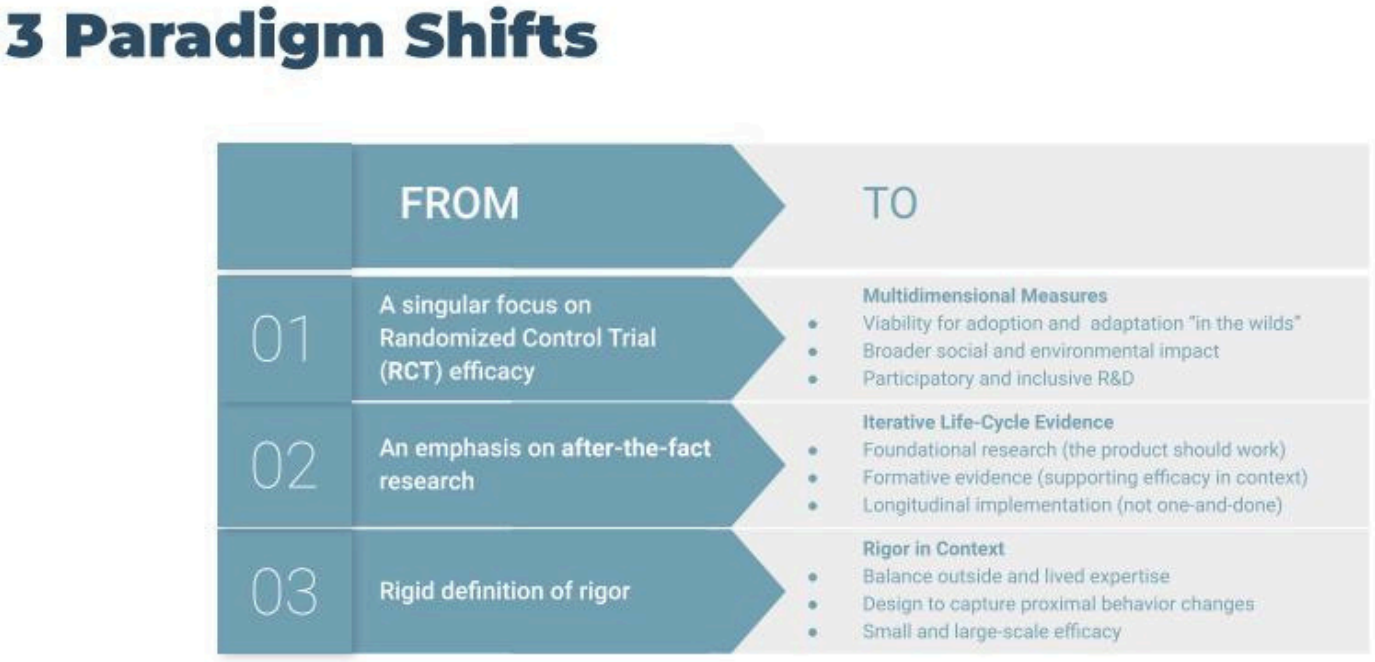

Figure 2: Summary of the paradigm shifts in the EdTech evidence field

Firstly, the recognition that impact is multidimensional rather than solely focused on efficacy. This shift acknowledges that impact encompasses various dimensions beyond just controlled experimental studies, including considerations such as cost- and pedagogical effectiveness, ethics, equity, and environmental effects.

Secondly, the understanding that impact occurs at all stages of the EdTech journey, and that the local adaptations over the duration of time influence the depth of impact in variable settings. This shift highlights the importance of considering the long-term implementation of EdTech solutions, recognizing that meaningful impact often requires sustained engagement over time.

Thirdly, the realisation that impact is a composite of multiple components, each with different relationships and weights depending on the rigour of the evidence. This shift emphasises the need for a nuanced approach to impact assessment, where different aspects of impact are weighted accordingly based on the strength of evidence supporting them.

These paradigmatic shifts complement each other and signify an evolving understanding of impact, not only within EdTech but also across the broader field of education. Namely, these shifts imply that achieving impact in EdTech necessitates not only effective products but also careful attention to the learning process and the individuals involved, particularly during implementation. It also means that attention to EdTech impact measurement is crucial at every stage of development, from conception to scaling across various educational settings. Balancing the weight of evidence ensures fair impact scoring and acknowledges the varying degrees of assumptions made with different measures and types of data.

Therefore, in proposing a new, more holistic, understanding of impact, we need to ensure that the criteria for different impact dimensions are balanced across products, processes, and people. This means not only testing products but also evaluating the educational process of contextualising the tools in different classrooms, replicating the use in different settings, and fully estimating whether they are worth teachers' time and schools' ' budgets.

Paradigm shift nr.1 : Impact is multidimensional

Moving beyond efficacy

The US evidence discourse has been led by the ESSA Evidence standards that propose a hierarchical understanding of evidence, with preference given to quantitative measures. This model emphasises experimental design as the leading research method for determining what works, with randomised controlled trials seen as the golden standard.

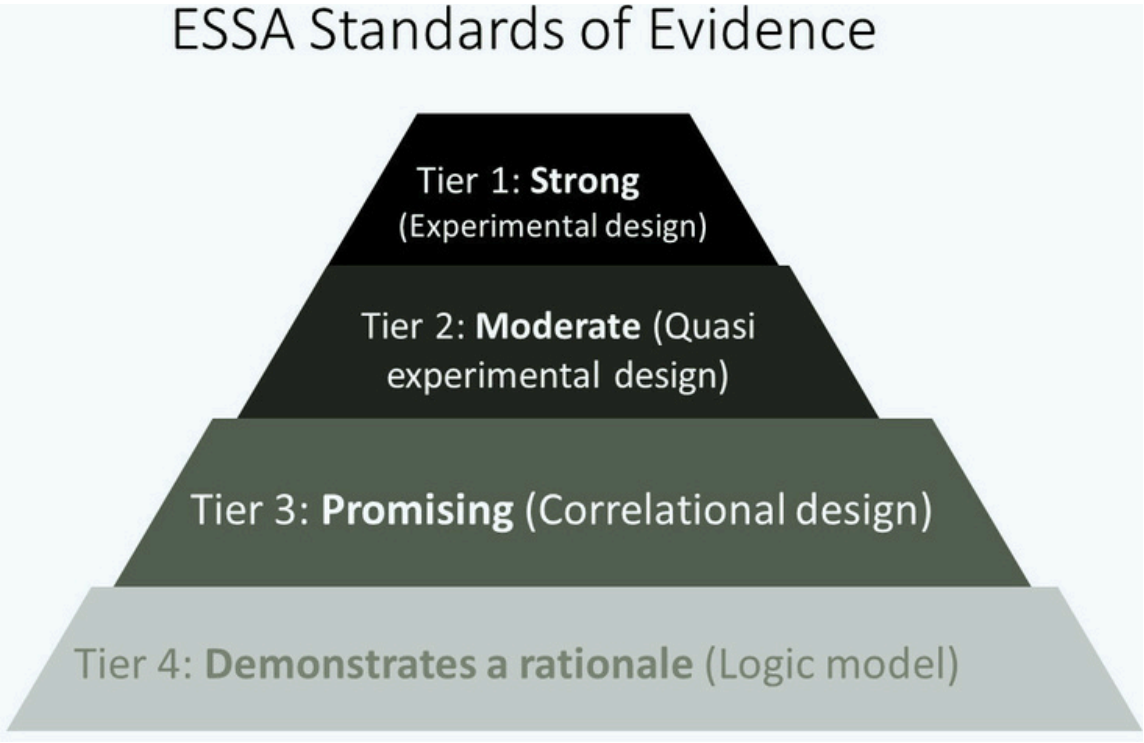

Figure 3: ESSA Tiers of Evidence (simplifed, based on https://tech.ed.gov/essa/

This paradigm offers a clear and consistent way for EdTech vendors to build their solutions and measure outcomes over time. At the same time, it neglects other types of research and evidence that matter in education. Notably, it reduces the evidence discussion to efficacy, which is an important element of evidence but insufficient for effectiveness (efficacy at scale) in a highly variable world. Furthermore, the importance of replication, teachers' insights and cost-effectiveness are evidence criteria that need to be considered in addition to efficacy.

Importance of replication

Andrew Coulson, Chief Data Science Officer at the MINDCET institute, emphasizes the need for shifting the narrative around EdTech research (Coulson, 2019), arguing that multiple field studies provide more valid insights than single "gold standard" experiments when assessing program impact. Coulson criticizes the reliance on the "one good study" paradigm, which has flourished with ESSA's emphasis on single RCTs as the proof that an EdTech tool works.

RCT studies, he notes, can take years to complete, often resulting in outdated evaluations that don't account for ongoing product revisions and improvements. Additionally, due to the complexity of full experimental roll-outs, RCTs often involve only one school district, which may not represent others' unique characteristics and needs. Furthermore, results from RCTs are specific to one assessment, which may not align with different state tests or changes within a state's testing over time. Also, RCTs usually focus on a specific grade level band, making it difficult to generalize findings across all grades, which vary significantly in content, teaching methods, student ages, and assessments.

Subsequently, the MINDCet Institute has proposed a framework for evaluating program effectiveness. This framework advocates for "a high volume of effectiveness studies as the future of a healthy market for product information in education" and emphasises the importance of repeatability, sample size, variety of assessment instruments (with more being better for higher ranking), range of grades, diversity of schools, study controls, independence of third-party studies, and product relevance.

The focus on replicability is an intentional step towards moving away from the understanding that "one good study" is enough evidence that a tool works. Reliable results, evidenced by replication from multiple studies can better support the notion that evaluations remain relevant and effective across diverse educational settings, ultimately leading to better educational outcomes. This understanding is embedded in the notion of an “evidence portfolio”, common to comprehensive education evaluations.

However, replicability overly implies the same intervention having the same impact under the same conditions in multiple contexts. What the learning sciences emphasise is research that reveals what works for whom under which conditions. Individuals (including students and teachers) and contexts (including classroom, school, and community cultures and implementation models) are highly variable in the real world. We need a portfolio of quantitative and qualitative assessments to understand the nuance of effective implementation.

A focus on marginalised students and equitable opportunities require a mix of methods that provides actionable insights into what works for whom. A recent analysis by Laurence Holt (2024), for example, found that industry efficacy studies typically restricted their analyses to students who had used the intervention programs with fidelity. That restriction, Holt concluded, left out a whopping 95% of participating students. The studies and their results didn’t represent the reality for most children and teachers and didn’t inform paths to improvement.

Co-design approaches

Various co-design approaches exist, building on the tradition of participatory research, that integrate users' perspectives into the development and scale process of EdTech from the start. So that user input is truly integrated into the full process of technology design, it needs to genuinely value the insights of those living in the impact ecosystem. This means collaborating closely with the intended users and proactively addressing the historical biases that have marginalised certain groups in EdTech.

Co-design, or participatory design, challenges the status quo by urging community organisations, researchers, and funders to recognize their own biases and assumptions. Co-design approaches centre equity in educational research and development (R&D) and the voices of users (learners or teachers) in the full R&D process. Angevine et al. (2019) promote a model called Inclusive Innovation, which redefines who holds power, makes decisions, and takes risks in education R&D. This model emphasises involving underrepresented groups from the very beginning, transforming their roles into leaders, active participants, and primary beneficiaries of research. Inclusive Innovation's true strength lies in not merely inviting diverse voices but positioning them at the core of the innovation process: "Inclusive Innovation is that it doesn't just invite underrepresented voices and perspectives into the innovation ecosystem; it places them at the center of it" (Angevine et al., 2019, p.3). (See the AERDF initiative and its inclusive Research and Development programme for an applied example of this approach.)

The Global Edtech Testbed Network is a collaborative effort aimed at gathering teacher-led interventions to test EdTech tools in real classroom settings through the perspective of “testbeds”.In testbeds, teachers play a central role in evaluating EdTech through rubrics and evaluations, with their perspectives strongly considered in assessing the value of each EdTech solution. The initiative operates on the principle that innovations should originate from classrooms, emphasizing small-scale iterations and implementation studies as essential steps in determining effectiveness. Rooted in the philosophical underpinning of educational research, this approach favors co-design and participatory methods over hierarchical and linear models of learning. It has yielded a substantial body of literature highlighting various learning patterns through qualitative approaches, identifying teacher and student preferences. Through iterative cycles of product co-design and pedagogical insights, this initiative aims to shift EdTech towards a paradigm of "pedtech," (Aubrey-Smith & Twining, 2023), where pedagogy plays a central role in shaping EdTech design.

Viability Cost-effectiveness

Viability refers to the extent to which an EdTech intervention is adoptable, affordable, and adaptable. What these terms mean, and the concept of viability, varies with perspective: from the perspective of low-income, high-need settings, cost-effectiveness is key. The key question to ask in low-income settings is: is the intervention worth it? The Global Education Evidence Advisory Panel (GEEAP), established under the auspices of the World Bank and UNESCO, helps make that judgement. GEEAP is an independent, interdisciplinary group of education experts, predominantly from academic backgrounds, convened to develop and provide guidance on cost-effective EdTech solutions that demonstrably improve learning in low- and middle-income countries. The group has produced a comprehensive “Smart Buys” report, offering evidence-based recommendations on which interventions are effective and which should be avoided, with cost-effectiveness as a key criterion. GEEAP regularly evaluates interventions and ranks them as 'Bad Buys, Good Buys and Smart Buys’.

The top-ranking “Smart Buys” are EdTech interventions that include supporting teachers with structured pedagogy programs, targeting teaching instruction based on learning level rather than grade (both in and out of school), and implementing successful early childhood programs, such as parent-directed early childhood stimulation for ages 0 to 36 months and quality pre-primary education for ages 3 to 5. Key to these considerations is the economic return on investment by a procurement team based on the implementation outcomes.

Although the framework/validation rubric for evaluating cost-effectiveness according to the smart buys is not publicly available, the group's work offers valuable insights into how to prioritise cost-effectiveness as a primary decision-making criterion for determining the effectiveness of EdTech solutions.

As for viability from the perspective of EdTech companies, the focus is on market opportunity. Is there demand, and is there funding? EdTech ventures perform market surveys of educator needs. They monitor legislative requirements and funding related to education. Pursuing equity goals may be a high priority for a company, but there needs to be a path for entities to adopt, pay for, and implement the intervention.

Broader implications

What are the broader implications of using EdTech? Like any other industry, EdTech has costs that have broader ethical and environmental considerations. For example, the use of large language models (LLMs) in educational tools is energy-intensive, raising environmental concerns. Increased screen time might reduce opportunities for in-person interactions. There’s also a social cost to treating bots as if they were human, and privacy may be compromised in the quest for personalised learning, especially when AI-driven tools, like chatbots, are involved. Beyond assessing the success of specific EdTech interventions, we need to consider these broader issues, including their efficacy, effectiveness, and impact on equity.

To connect to these broader implications while keeping educational impact at the heart of the evaluations, Kucirkova (2023) proposed the 5Es framework —comprising efficacy, effectiveness, ethics, equity, and environment. The 5Es can be understood as an umbrella framework for evaluating the impact of education and EdTech: efficacy is concerned with the technology's influence on learning, social dynamics, and economic outcomes, while effectiveness gauges EdTech ability to address cost, pedagogical approaches, and local infrastructure challenges. Equity examines the extent to which EdTech centre marginalised groups, while ethical considerations encompass transparent, safe and responsible methods for data processing. Environmental impact is concerned with the EdTech's impact on local and wider environment, including planetary health.

The 5Es framework aligns with the 17 Sustainable Development Goals (SDGs), with each vertical emphasizing a group of specific goals. For instance, efficacy corresponds to SDG 3 (promoting health and wellbeing), SDG 4 (ensuring quality education), and SDG 8 (supporting economic growth).

While the 5Es separate the individual impact dimensions into distinct pathways for evaluating impact, with separate objective metrics for each, the individual dimensions are interconnected, with each "E" influencing others. Thus, while certain impacts may take precedence in certain contexts, their relevance extends across various areas and the primary objective is to maintain a balanced approach to achieving progress across all impact domains.

The 5Es framework served as the foundation for developing a multidimensional impact index (Kucirkova & Cermakova, accepted), which incorporated specific indicators for each level. This initiative was part of the Multidimensional Impact Index project, which convened researchers specializing in the respective strands of the framework. These research groups systematically reviewed relevant frameworks and literature for each vertical, including industry-specific evaluation criteria, resulting in the publication of independent reports. Additionally, consultations with industry stakeholders and policymakers were conducted to refine individual indicators, contributing to the overarching evaluation framework. Ultimately, this evaluation framework was adopted by the International Council for Education in EdTech (ICEIE) to establish certification criteria for EdTech solutions.

The importance of viability

The various perspectives of members of the ecosystem can be woven into the dimensions of effectiveness, ethics, equity, and the environment. These dimensions matter theoretically, but how do they matter within the ecosystem? Viability is a critical dimension, which needs to be incorporated into all E dimensions. It prompts us to ask: how viable is the proposition that Edtech can achieve strong impact in all dimensions? Whose responsibility is it to support the process for this to happen? Is it realistic to expect EdTech companies to have positive impact on all dimensions?

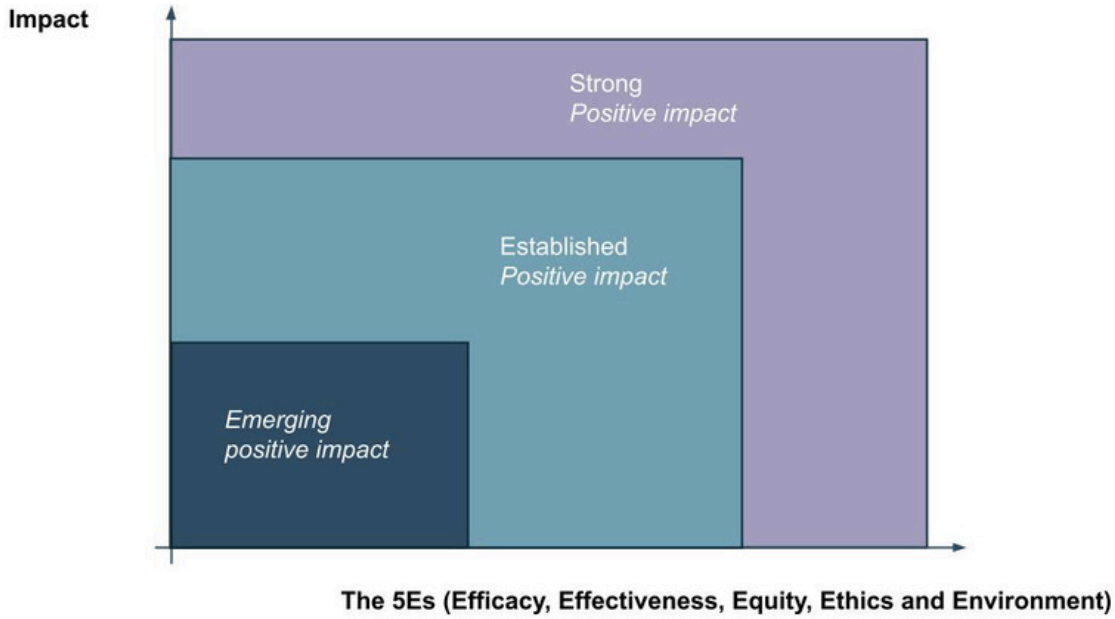

Figure 4: Graph that captures how impact and Evidence (as in terms of the five types of evidence) increase from emerging to strong

Paradigm shift nr.2: Iterative Evidence Over Time

As noted above, evidence of impact is all too often an after-the-fact consideration. Did the intervention work? This question is complex on its own, and it comes too late to inform design and implementation. Instead of treating evaluations of efficacy as something akin to a summative assessment, we recommend a more iterative evidence paradigm throughout the entire EdTech product lifecycle. We want interventions to succeed, and evidence can inform and improve the chances of the desired impact.

Research and evidence need to be infused into EdTech throughout the process, from problem definition to long-term efficacy at scale. In addition, working with and through technology provides opportunities to make processes, as well as outcomes, visible.

We get to see how learners achieve their goals. That potential window into behaviour provides critical information for capacity building, for helping children become more effective learners and teachers more effective facilitators of that learning.We suggest three guiding questions for EdTech founders, funders, and researchers to consider in this process.

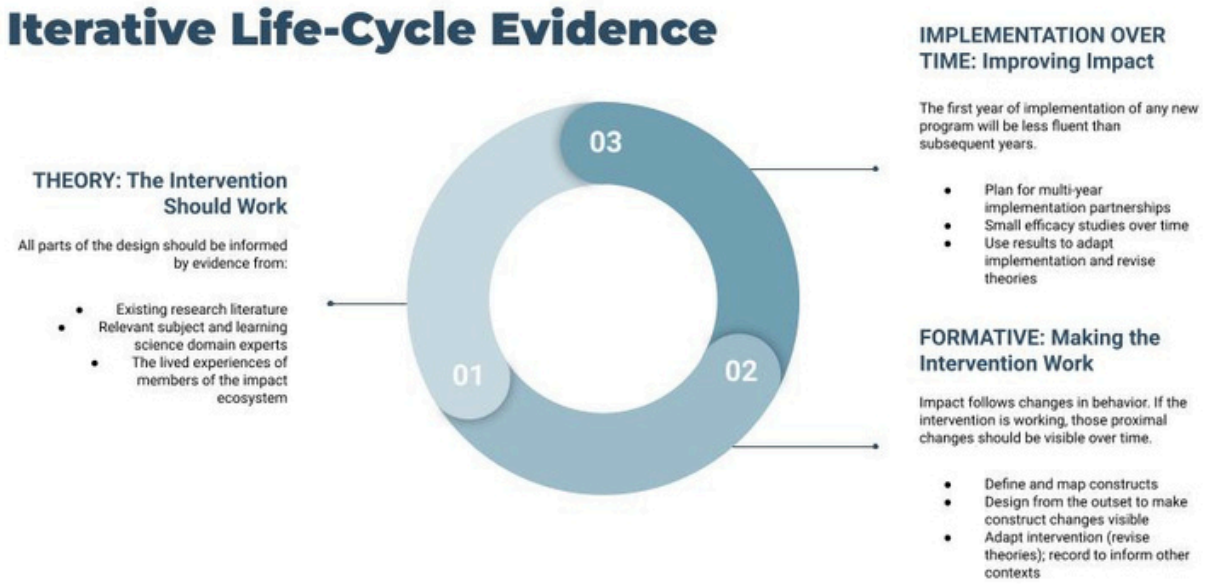

Figure 5: Depiction of the iterative cycle of generating evidence from theory through formative and implementation research

Question 1: Why should the intervention work?

The base level of ESSA evidence requires a logic model for the implementation of an educational intervention. We want to elevate the research and theories that instil confidence in that logic model, what historically has been called research-based design, but that today reflects a more inclusive approach, both in terms of user involvement and holistic measures of impact.

EdTech developers, whether academics building prototypes or commercial ventures launching businesses, should identify and interrogate their assumptions about the problem they hope to address. They should review existing research literature, consult with domain experts, and engage with members of the impact ecosystem to determine addressable causes of the educational inequity. They should articulate what constructs - anything from knowledge and skills to dispositions and behaviours - need to change and for whom. And then they must propose theories of action for causing those changes, along with a rationale for why each theory should be successful. If (user action)...then (construct change)...so that (observable impact)...because (underlying research)....

A given intervention will have multiple theories of action. Indeed, every design choice should have an underlying rationale, and many, if not all, of those theories of action should be testable. Does, for instance, a particular type of growth mindset feedback promote student persistence? Sometimes evidence supporting an individual theory exists in the research literature, as it does for different types of feedback. Sometimes it requires a new prototype for validation. The more compelling the underlying research and evidence, the more confidence in the design that the intervention should work.

Further, the intervention design should anticipate and address ethical and environmental considerations, in light of equity and sustainability goals. Will the intervention be accessible - in terms of cost, infrastructure demands, usability, and so on - to the target populations? Will user information be treated securely and fairly? Will the intervention have potential spillover social, cultural, or environmental consequences that might be adverse to local or other populations? The research informing these questions may come from economics, philosophy, and other non-pedagogical sources. When it comes to the lives of children, responsible EdTech development shouldn't rush to break things and worry about the clean up later.

Question 2: Is the intervention working?

Formative assessments are meant to capture student progress toward performance outcomes, typically to inform changes in instruction. Well-designed formative measures can reveal levels of student mastery or understanding. They provide insight into how effectively an intervention is producing outcomes, but it often fails to expose the behavioural shifts leading to those outcomes.If students and teachers keep doing the same things, they will get the same results. Impact follows changes in behaviour. Those proximal behavioural changes should be noticeable, providing evidence of if and how well an intervention is working to drive the desired outcomes.

Answering Question 1 will help in collecting data for Question 2. Testing prototypes of individual theories of action requires defining what counts as evidence of change. Here, multiple and diverse forms of evidence are important, targeting not only easily measurable outcomes but also "life outcomes" that matter but might not have standardized measurement tools. How, for instance, do we know that student persistence is increasing ahead of the improved performance outcomes that will follow from the behaviour shift? Designing interventions to make visible leading behavioural indicators of promised outcomes is challenging (see below under Rigor of Evidence), but it contributes to an understanding of what changes for which individuals are happening as planned and what adjustments might be needed to get the process on track. And the needs and adjustments will likely not be the same in each implementation context. As we know, communities, schools, and classrooms are complex and variable worlds.

Question 3: Did the intervention work?

Being rigorous about Questions 1 and 2 should, theoretically, improve the chances of a positive answer to Question 3. More importantly, this more holistic approach to research and evidence should provide more insightful data about how well the intervention worked, for whom, and under what conditions. It can also inform how to make the intervention work better over time. Just as we extend the consideration of research and evidence over the whole product life cycle, from conception to implementation, we also want to extend the notion of implementation as an ongoing, iterative process, both within a setting and across contexts.

Taking these questions into account, we can progress our understanding of impact and time relationships.

The Impact x Time diagram below represents a within-context view of impact over time. Long-term implementation typically devises strategies for sustained adoption, gradually reaching mindset shifts.

There is a rich body of work aligned with this thinking, known as “research-based design” and encompassing design-based research, design-based implementation research, and translational science. The field of Design-Based Implementation Research (DBIR), in particular, recognizes the importance of evaluating the process of turning research-based practices and innovations into effective practices, a transition the biomedical world has come to call translational science. DBIR and translational science connect to the impact x time equation by prompting the question: “How well has the intervention worked so far?”

It takes time for an intervention to become fully embedded in a setting, and it is vital for including time considerations in questions of “impact”.

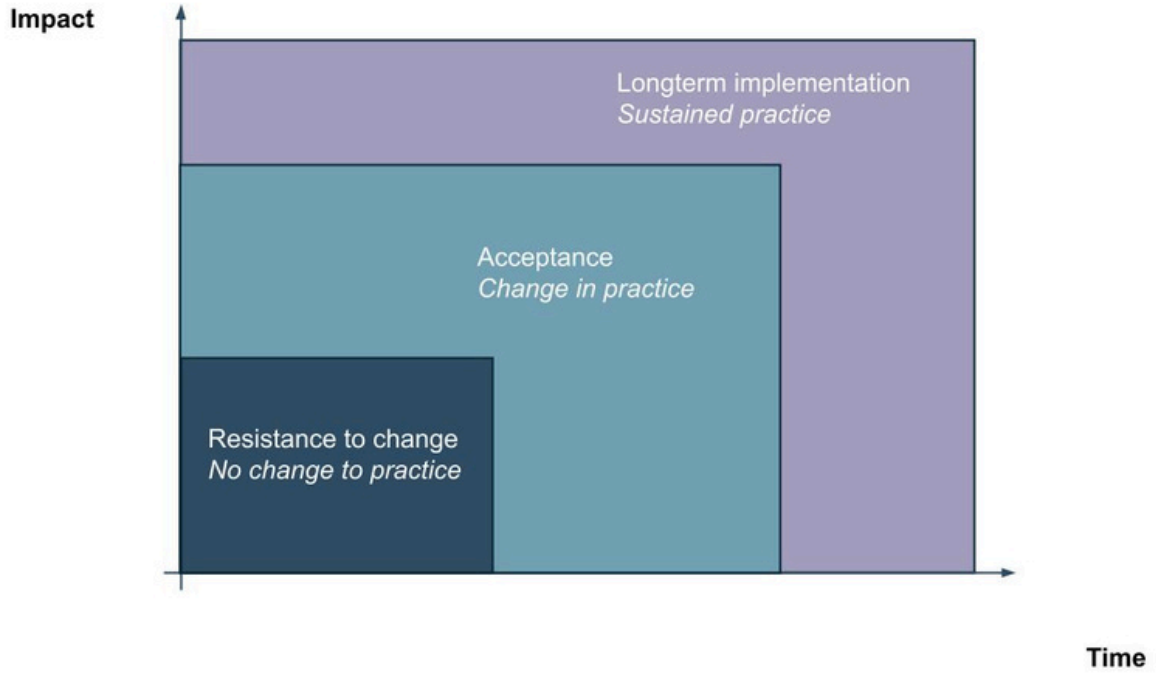

Figure 6: Graph depicting how impact and time relate to changes to practice, from minimal to sustained changes over time

In addition, to time, we need to address question of contexts to fully capture the reality of EdTech interventions. The Impact x Scale diagram captures this additional, r critical, dimension of impact: effectiveness across settings. The efficacy of early prototypes and versions of an intervention fuel a desire to scale the impact. Translating impact from one setting to another, however, is challenging. This is not about replicating the same approach across diverse settings to see whether it “truly works”. Rather, it is about aligning the intervention with the specific conditions of different educational contexts. This alignment is at the heart of ethical implementation science. The diagram shows the typical reduction in impact that follows scale.

Designing from the outset to capture data related to processes by setting and subpopulations in the ecosystem, we argue, will help to identify ways to improve impact as an intervention scales. It could also provide insights to improve impact among efficacy and early adopter settings. As emphasized by the equity dimension earlier, the marginalized groups in those sites should be central in future implementations. Using evidence to inform ongoing implementation over time can drive changes in prior implementations.

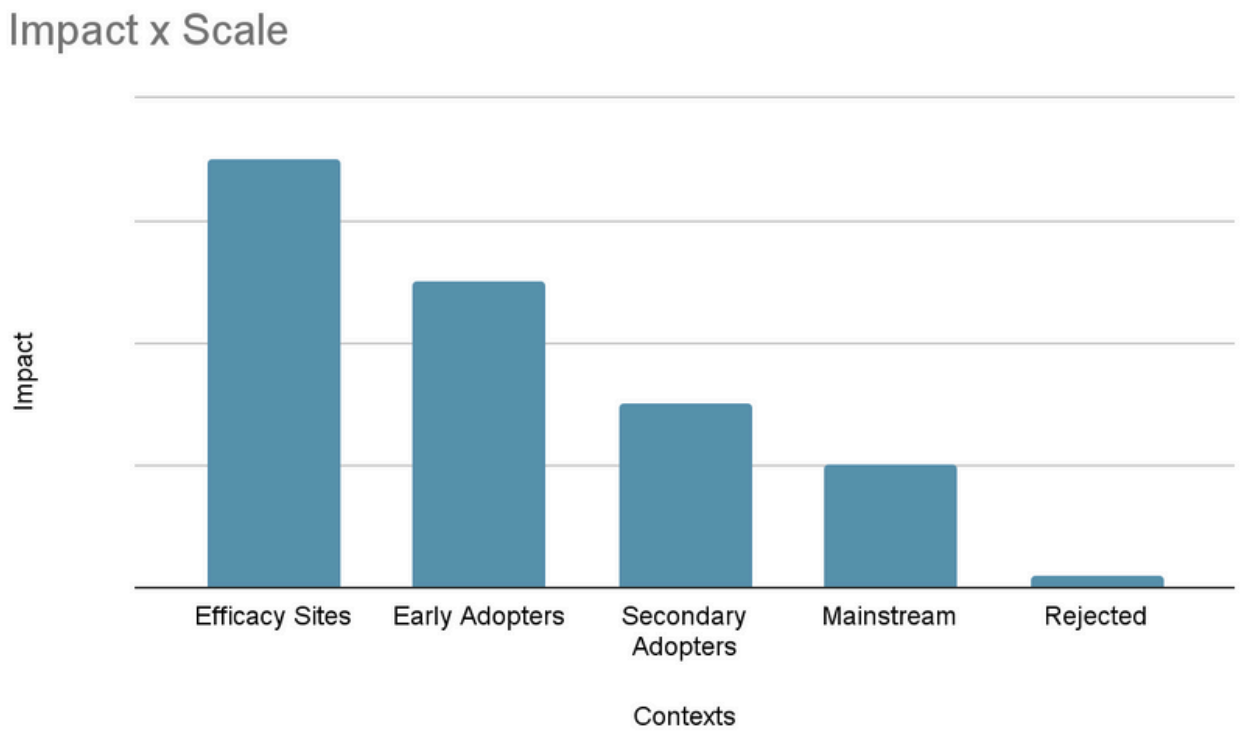

Figure 7: Graph depicting how impact intersects with scale across multiple contexts

Paradigm shift nr.3: Rigor in Context

Our third paradigm shift centers on enhancing the quality and rigour of evidence. ESSA places randomized control trials (RCTs) at the pinnacle of evidence rigour for EdTech. We agree that RCTs do represent a rigorous analysis of an intervention's efficacy, but RCTs alone are insufficient to inform effectiveness, or efficacy at scale, across varied settings. We argue for a holistic portfolio of evidence that is iterative and spans the lifecycle of product ideation through long-term implementation. The rigor of evidence should be appropriate to the context in which it is used.

Given the capabilities of technology to collect and analyze data, including new 3rd party research studies, in real time, we urge an interactive framework where evidence is constantly being collected and used to inform product design, redesign, and implementation. This framework, which we outline in more detail below, requires a shift across the ecosystem, among developers and educators, as well as researchers. The general process can be reflected in this simple diagram.

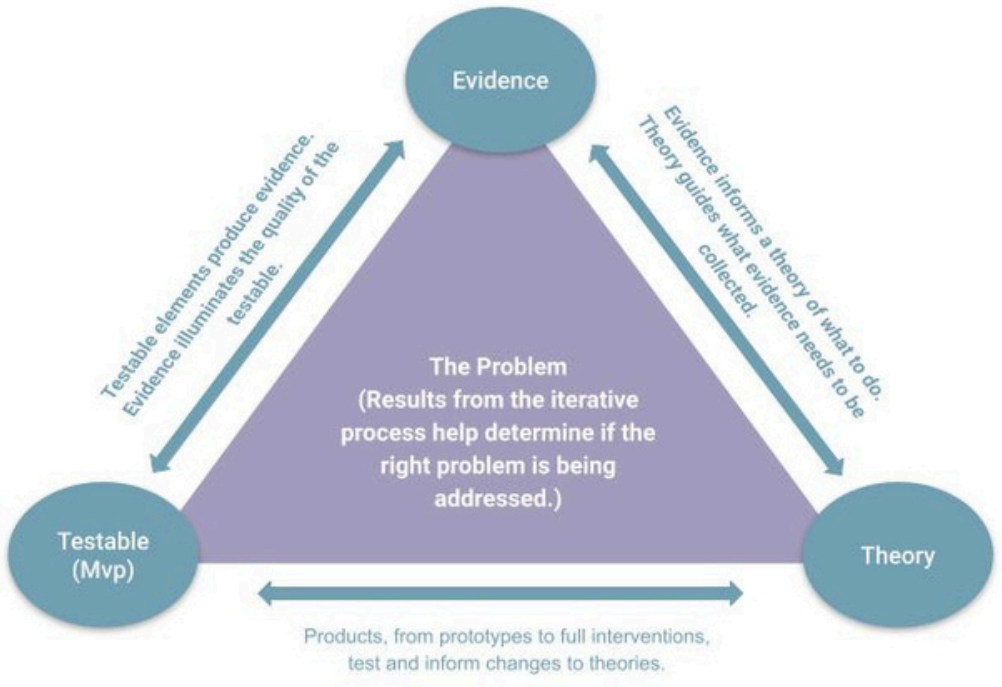

Figure 8: Model capturing the iterative cycle of influences from evidence to theory and testable MVP

The implementation of an EdTech product is, in essence, a test of its Theory of Change, typically captured in a conceptually-rich Logic Model. The theory of change and logic model provide the evidence for the claim that "The product should work because…." The evidence from an implementation study may reveal that the theory needs revision, and/or it could reveal problems with the way the theory was manifested in the product. Or, maybe the result was heavily influenced by the way it was implemented. It is also possible that the product was trying to solve the wrong problem. It is complicated. Indeed, carrying out this test when a product is finished, encapsulating potentially hundreds of individual theories related to every design decision, is usually too late to reveal specific elements that work well and those that don't, for whom and under what conditions. Implementing a rigorous testing and evidence collection process from the outset and testing individual assumptions in relation to specific design features is a good way to probe the effects of specific mechanisms. Then, designing a product that anticipates ongoing iteration, we suggest, will ultimately lead to products whose effectiveness increases over time and informs future development that can expand the effectiveness to diverse groups of users.

Q1 Rigour: Informing Why It Should Work

These earliest product stages have no prototype interventions to test for efficacy, but they have plenty of assumptions, or theories, that need to be checked. Many educational products are developed to solve a problem. EdTech, for instance, often highlights the shortfalls of one-size-fits-all classroom instruction. Many students fail to achieve, it is often argued, because teachers can't meet the individual needs of each student (Dockterman, 2018).

Technology-based products, innovators argue, can leverage AI-driven adaptivity to adjust content to match the level and even interest of each student. These theories about cause and solution have driven the development of generations of EdTech programs that, generally, have yet to make a significant impact on student performance (UNESCO GEM Report, 2023)]. Instead of solving a problem, the EdTech may have contributed to the problem of focusing on products and standardised ways of demonstrating knowledge rather than the processes of learning. In that regard, our question is: How strong was the evidence in each case that the intervention should have worked in the first place?

Bringing rigour to this question requires engaging deeply with at least threesources of evidence around the basic assumptions related to the problem, its causes, and the potential changes that can drive impact. The differing perspectives of these three sources - existing research literature, domain experts, and members of the impact ecosystem - contributes to the definition of rigour at this stage. Existing studies from the relevant learning sciences - including cognitive psychology, neuroscience, behavioural psychology, and so on - provide an academic lens that is particularly relevant to the cognitive, psychological, and sociological understanding of problems and causes. Learning sciences and learning engineering can illuminate learning mechanisms, efficacious instructional approaches, and both biological and socio-economic issues that foster effective educational processes.

Experts from the relevant content and learning science domains are helpful in adjudicating tensions typically found in the literature. Neurobiologists study learning and teaching in ways that are different from, for instance, psychologists or economists. Insights from an fMRI study may reveal differences, say, in the way neurotypical and atypical children process aspects of language or mathematics. Psychologists may report on how teacher feedback impacts student academic identity, and economists may report on the use of behavioural nudges or financial incentives for motivation. Interrogating various experts can help uncover the core problems, among many, that a particular new intervention should address. Outside expertise, however, must be balanced with the internal expertise of those living within the ecosystem any new intervention intends to infiltrate.

The Ecosystem perspective

A K-12 ecosystem can be quite complex. The table below captures the beginnings of who might be included. Clearly the children and their teachers are part of the ecosystem. Administrators, parents, and IT personnel could also be involved in the acquisition and implementation of a learning intervention. The ecosystem could be further extended to capture school board members, special education or content area specialists, and even vendors who supply infrastructure and curriculum materials.

The various members of the ecosystem have different values. A child, for instance, may prize social acceptance. A teacher may seek a manageable instructional environment. An administrator could crave improved test scores, while parents and IT personnel may be concerned about wellbeing and safety. There's also variability within each column. Not all children and teachers are alike. A rigorous analysis to define the problem, identify the addressable causes, and generate potential solutions will include rich engagement across and within the ecosystem in an inclusive, dialogic, process.

Table 1: Key stakeholders and concepts in the K12-Impact Ecosystem

K-12 Impact Ecosystem						
	Child	Teacher	Administrator	Parent	IT Director	Others?
Educational Values						
Anticipated Variability						
Key Constructs						
Evidence of Change						

A convincing research foundation and logic model for why an intervention should work will reflect multiple sources of evidence and an iterative process that articulates, challenges, and tests assumptions about what needs to change, for whom, and how. It should have a plan for adding evidence of engagement with all the stakeholders around the four areas of impact, in all the cells illustrated in the table.

Rigor Q2: Collecting Evidence That It IS Working

Outcomes follow long-term changes in attitudes or behaviour, and such changes in the process of learning and teaching need time to take hold. A robust logic model or theory of change will describe those changes. Teachers, for example, will not only receive professional development, they will adjust the way they instruct. Students will not only hear growth mindset feedback, they will take more academic risks and more actively persist in the face of academic challenges. These kinds of shifts in behaviour precede and lead to better outcomes. Behaviour is measurable and observable and both quantitative and qualitative methods are key for a full understanding. Making evidence of behaviour changes visible offers a window into how well an intervention is working and can suggest modifications to improve the chances of impact.

Monitoring and evaluating various levels of impact at various time intervals is an essential part of collecting evidence. Seeking evidence of behaviour is different from assessing interim progress toward long-term outcomes, which is also valuable. Progress monitoring assessments are common and typically administered multiple times a year. These assessments reveal changes in outcome, but they provide little insight into what changes are or aren't happening that contribute to the outcome results. They tend to be binary: Yes, students are improving as expected; No, students aren't improving as expected. We again argue for a more holistic and iterative approach to evidence in this process.

Evidence-Centered Design is one approach to assessment that can be readily adapted in service of learning to support this evidence collection effort (Arieli-Attali, et.al., 2019). Members of the design ecosystem - learning designers/innovators, educators, and researchers - need to work together to define the constructs, and their respective ontologies, or construct maps, that need to change for the various members of the impact ecosystem. The ontologies will describe the elements and progressions of evidence. What, for example, does it look like as a teacher grows increasingly proficient with a new instructional approach? What is the difference, say, between a student who is a little persistent versus one who is mostly persistent versus one who is very persistent? Describing those variations in behaviour guides the design of mechanisms to expose and observe them.

Surveys and quizzes are relatively easy to design and administer and many EdTech tools on the market contain embedded surveys and quizzes to collect self-reported data on behaviour or attitude change. We suggest that direct observations of behaviour can usefully supplement surveys and quizzes with more actionable and user-empowering information. When teachers, instructional support personnel, student peers, and the learners themselves are actively noticing desirable behaviours, they have opportunities to adjust in the moment or near-moment. Tracking the observations and adjustments provides a much richer view of what changes are happening for whom under what conditions. With guidance from researchers, developers can be intentionally designing for such self-monitoring and observations. Building opportunities for observations into the product and its implementation (rather than always seeing observations as a research add-on), is a way to collect evidence in a collectively instructional manner.

Q3 Rigor: Did It Work (and how can it work better at scale)?

Our third paradigm shift centers on enhancing the quality and rigor of outcome evidence. As active researchers, we aim to define and cultivate rigorous evidence within and across our organizing principles, focusing on impact along the vertical axis and settings, or context, along the horizontal axis.

When considering different contexts, we initially examine smaller contexts with limited impact. As implementation contexts expand, such as when technology scales across classrooms and school districts, the impact becomes more significant. However, ensuring this impact is positive requires evaluating it at scale, aligning evaluation methods with the complexity of determining impact.

While a small-scale study may provide valuable insights, it lacks the capacity to assess impact across diverse contexts comprehensively. Therefore, the weight of evidence becomes crucial in evaluating technologies. Rather than favoring a hierarchy between quantitative and qualitative methods, learning sciences emphasize that both are essential in EdTech evaluations, and both can increase the depth of an analysis (Kucirkova, Brod & Gaab, 2023). Rigorous application of both methods is essential, necessitating adequate resources such as time for analysis and personnel for data collection. These understandings of rigor are well-established in research, (even though equitable distribution of resources across the academic field remains a pertinent consideration), and they should be applied to EdTech evidence-building too.

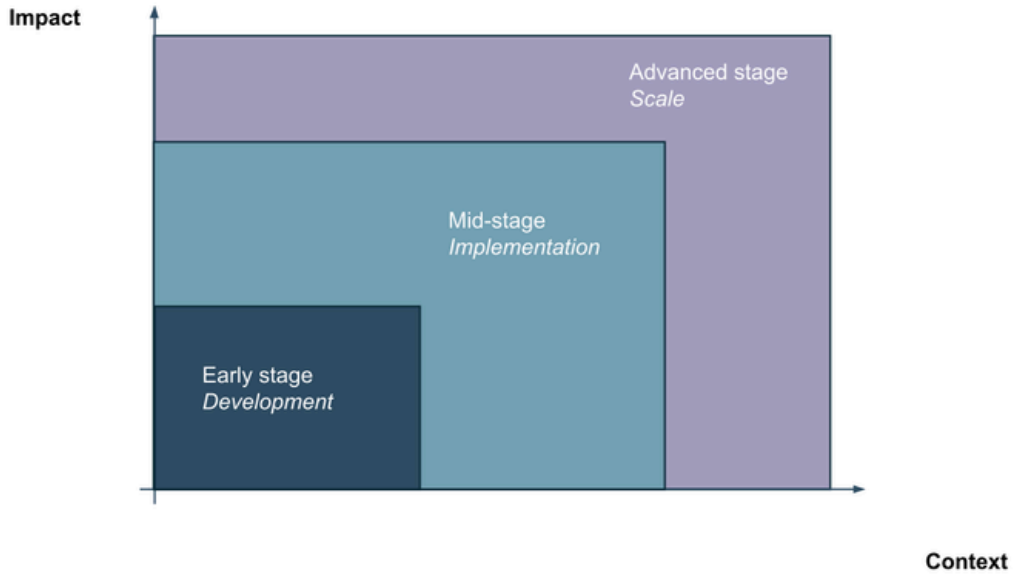

Figure 9: Graph depicting the relationship between Impact and Context influencing the individual stages of EdTech

The various stages of EdTech implementation require distinct investments from different members of the ecosystem. In smaller-scale contexts, product teams collaborate closely with developers, and small-scale studies involve active participation from teachers. However, as the technology scales up, marketing and sales teams become more detached from the product development process, and teachers are often positioned primarily as consumers rather than active evaluators or feedback contributors. At significant scales, such as with big tech, teachers predominantly serve as consumers, with limited opportunities for active contribution to design. In other words, in smaller-scale contexts, teachers may have more direct involvement in shaping the implementation process, whereas in larger-scale contexts, their role may be more passive, primarily focused on utilizing the technology rather than influencing its development or evaluation.

Conclusion

In this working paper, we proposed three paradigm shifts and offered some conceptual tools for thinking about the impact of EdTech solutions holistically and with attention to efficacy as well as effectiveness, ethics, equity and environment. The ideas and suggested solutions here call for a new kind of alignment and collaboration among all members of the EdTech ecosystem.

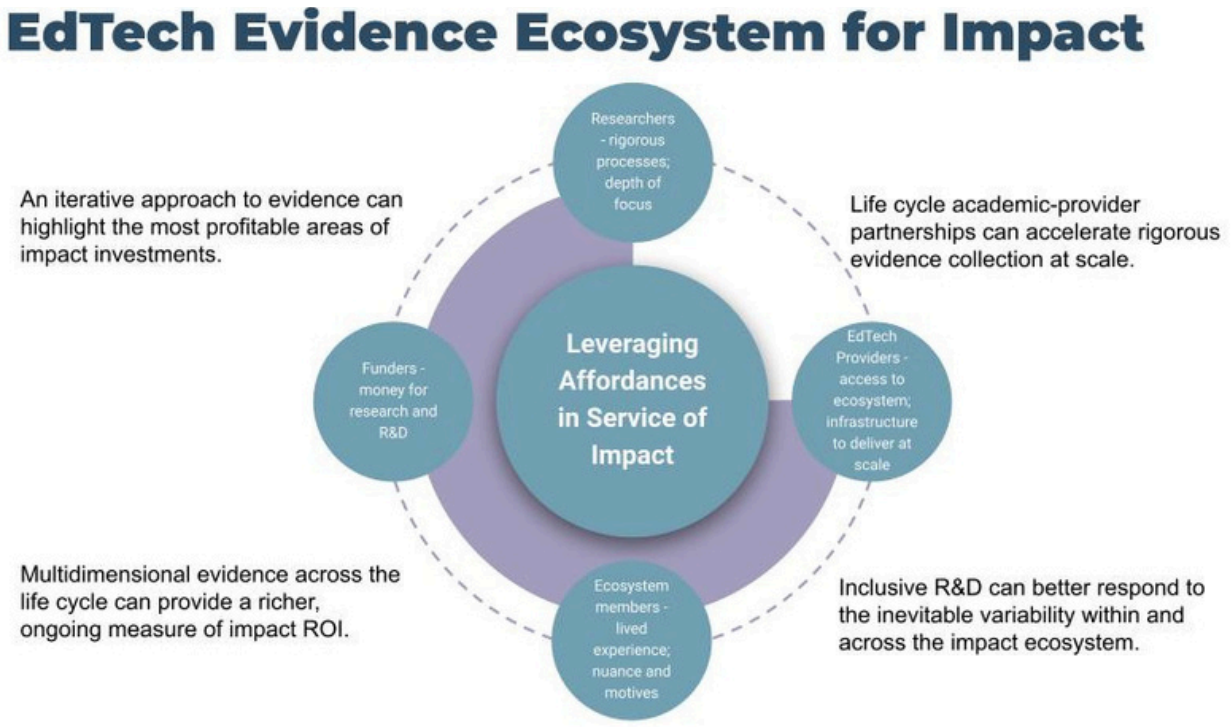

Figure 10: Visualisation of the EdTech Impact Ecosystem

Funders, publishers/innovators, educators, and researchers all have critical affordances that need to work in alignment to generate impact that can be scaled and sustained. Researchers have the wherewithal to define and measure impact over time and settings. EdTech publishers and founders have the infrastructure to access and support variable contexts. Educators, including children and parents, have the lived experience and are the audiences we serve. Funders - including philanthropies, governments/policymakers, and impact investors - direct the action through the flow of money. A stronger evidence story enables us to track both financial and impact returns on our investments over time. Indeed, we want the financial return on investment to be in service of the educational return on investment and that can only be achieved with a systems-level perspective and a holistic approach to evidence.

We hope that our paper will serve as a catalyst for more nuanced thinking and conversation around impact in schools and capacity-building within the EdTech industry. We invite your comments to move this conversation forward.

References

Angevine, C., Cator, K., Liberman, B., Smith, K., & Young, V. (2019). Designing a Process for Inclusive Innovation: A Radical Commitment to Equity. Version 1.0. Digital Promise.

Arieli-Attali M, Ward S, Thomas J, Deonovic B, von Davier AA. The Expanded Evidence-Centered Design (e-ECD) for Learning and Assessment Systems: A Framework for Incorporating Learning Goals and Processes Within Assessment Design. Front Psychol. 2019 Apr 26;10:853. doi: 10.3389/fpsyg.2019.00853. PMID: 31105616; PMCID: PMC6498139.

Aubrey-Smith, F., & Twining, P. (2023). From EdTech to PedTech: Changing the Way We Think about Digital Technology. Taylor & Francis.

Coulson, A. (2019). Why You Need More Than "One Good Study" To Evaluate EdTech, Why You Need More Than "One Good Study" To Evaluate EdTech (mindresearch.org)

Dockterman, D. (2018). Insights from 200+ years of personalized learning. npj Science of Learning, 3(1), 15

Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., & Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. Teachers College Record, 115(14), 136-156.
Global Education Evidence Advisory Panel: https://www.worldbank.org/en/topic/teachingandlearning/brief/global-education-evidence-advisory-panel

Holt, L. (2024) The 5 Percent Problem: Online mathematics programs may benefit most the kids who need it least, Education Next, https://www.educationnext.org/5-percent-problem-online-mathematics-programs-may-benefit-most-kids-who-need-it-least/

Kucirkova, N., Brod, G., & Gaab, N. (2023). Applying the science of learning to EdTech evidence evaluations using the EdTech Evidence Evaluation Routine (EVER). npj Science of Learning, 8(1), 35.

Macgilchrist, F., Potter, J., & Williamson, B. (2024). Challenging the inequitable impacts of edtech. Learning, Media and Technology, 1-4.

OECD (2018). The Future of Education and Skills: Education 2030. The Future We Want,Available online from: https://www.oecd.org/education/2030/E2030%2520Position%2520Paper%2520(05.04.2018).pdf

UNESCO (2023) Technology in education A tool on whose terms?, GEM Report, Available online from: https://www.unesco.org/gem-report/en/technology

Suggested citation:

Kucirkova,N. & Dockterman, D. (2024). Towards a holistic understanding of evidence: A working paper. *International Centre for EdTech Impact,* Creative Commons Attribution 4.0 International. DOI: 10.13140/RG.2.2.19241.81767