

Measuring the Impact of Emerging Technologies in Education: A Pragmatic Approach

Mutlu Cukurova and Rose Luckin
UCL Knowledge Lab, University College London
London

Abstract: *The evaluation of emerging technologies is important for their impacts to be effectively integrated into learning and teaching settings to bring the best benefit to learners and teachers. Educators, learners, parents and policymakers alike, therefore, need reliable methodologies for evaluating the effectiveness of such emerging technologies. However, the impact evaluations of technology in education are challenging. This challenge is more significant for emerging technologies, as change is the essence of emerging educational technologies. Therefore, the value of traditional impact evaluations in education require being reconsidered within this context. Here, we present a pragmatic approach to measuring the impact of emerging technologies in education which focuses on the suitability of the proposed evaluation methods and the types of evidence rather than on the hierarchy of these methods and evidence types. The approach has two main steps. First one is the creation of a clear theory of change to identify outcome measure(s) and assumptions that are behind the expected impact of the emerging technology intervention. Secondly, the identification of the type of evidence and methods to generate it that are the most appropriate for the current innovation stage of the emerging technology.*

1. Introduction

The impact of technology on learning and teaching is often at the forefront of demands, particularly from those who dictate the funding available to pay for technology within education systems. This is not an unreasonable request and there is merit in impact evaluations of educational technologies including emerging technologies. Nevertheless, as has been shown in numerous meta-level investigations (see for instance Cox et al., 2003) evaluation of the impact of technology on educational outcomes is a challenging task. This challenge is even greater when evaluating emerging technologies. This is at least partially due to the connotation that, in traditional impact evaluations, evidence regarding the impact of an intervention is considered as *a shield against change*. The generation of scientifically robust evidence about the impact of an educational technology can be therefore taken as a message for the stakeholders of this technology to standardise it and scale it up. *However, change is the essence of emerging technologies.* Measuring the impact on educational outcomes is certainly a necessity if we are to increase our confidence in our educational technologies for their potential to meet

their expected outcomes. However, particularly in the case of emerging technologies, the evidence generated from impact evaluations might rather be used to provide 'informed change' instead of a justification for standardized practice and manuals for implementations.

This fundamental difference between emerging technologies and more traditional educational interventions is important and it must be taken into account in the discussions about, and designs of studies to evaluate the impact of emerging technologies. In the context of emerging technologies, the value is in the careful consideration of different types of evidence as well as robust methods to generate them. In this chapter, we will focus on two key particularities of emerging technology interventions, and based on these, we will suggest a pragmatic approach to evaluate the impact of emerging technologies.

1.1 Defining Emerging Technology

When simply defined, particularly within the context of education, emerging technologies are those that have the potential to change the current state of affairs in education. Emerging technologies include but are not limited to virtual reality implementations (Merchant et al., 2014), augmented reality implementations (Dunleavy, & Dede, 2014), mobile learning devices (Crompton, Diane, & Gregory, 2017), physical computing tools (Katterfeldt et al., 2018), internet of things hardware with sensors (Cukurova et al., 2018), and technologies that allow collaborative learning at a great scale (Cress, Moskaliuk, & Jeong, 2016). Change is at the core of these technologies both because they evolve by time, but also arguably their *raison d'être* is to transform the learners' current experience. Therefore, the change is not only in the educational technology that is aimed to be evaluated, but also, in the learners' experience of the interventions through which the impact of the technology will be evaluated. When change is such fundamental to emerging technologies, how can we measure their impact in a robust and meaningful manner? Next, we will go through the notion of evidence-based education and how evidence is being generated to measure impact in more traditional educational settings with the purpose of presenting the (in)compatibility of these ideas to the nature of emerging technologies.

2. Evidence-informed Emerging Technologies in Education

The notion of impact and the demand to be shown that something really 'works' requires evidence of some kind of influence or effect of the educational technology on their users. In educational technologies, as well as in education broadly, often this evidence of effect is measured in terms of learner attainment (see for instance Cox, & Abbott, 2004). Measuring impact on student attainment might be considered as somewhat narrow, and increasingly there is recognition of the value of illustrating effects on a much wider range of outcomes. The previous edition of this book presents some of these examples such as those relating to emotional well-being (Ainley, Enger, & Searle, 2008) or motivation (Cox, 2008). Regardless of the type of outcome measured to show the impact, high-quality evidence that relates to the

potential of an educational technology to achieve its expected learning outcomes and the information regarding the process of how these learning outcomes can be achieved is key for impact measurements.

Although there are some concerns regarding the centralization of such evidence as the only source of input for decision making on which to base educational practice (see, for instance, Biesta, 2007), it is hard to argue against the potential value of evidence to inform and improve practice (Petty, 2009). Recently, interest in effectiveness evidence in educational practice and policy has increased globally (Buck & McGee, 2015; Greany & Maxwell, 2017). For instance, in the UK, the department for education invested a £125 million founding grant to Education Endowment Foundation in 2011 aiming to improve evidence-based practice in education. Department for International Development is preparing an approximate £20 million investment to create a global research hub to form a global 'what works' evidence hub to catalyse effective innovation in the education sector. This initiative aims to provide decision-makers with evidence to harness the transformative potential of educational technology – to deliver better learning outcomes for all (DfID, 2017). Similarly, in the USA, 2015 Every Student Succeeds Act encourages the use of an educational practice that meets evidence standards from experimental or quasi-experimental evaluation studies. Fundamentally, these initiatives point out that educational policies and practice should be chosen based on evidence on their effectiveness (Slavin, 2017).

Governmental initiatives as mentioned above, lead to discussions around impact and evidence in primary and secondary education practice with an emphasize on 'what works', suggesting that practice should implement teaching methods that are presented to be effective at maximizing learning outcomes by research (Hattie, 2008). However, despite the significant amount of academic interest and policy focus as well as the resources directed to educational technology's impact, it is still a significant challenge to generate and identify robust evidence on the effects of technology in education, particularly for emerging technologies.

After decades of research in the field of technology in education, there are still significant questions waiting to be answered that relate to the impact of technology on students' learning (Cox, 2008). There are various reasons behind this lack of robust and reliable evidence regarding the effects of technology in education.

As categorized by Cox and Marshall (2007), these include issues that relate to different assumptions about how learners think, the broad range of technology types and their varied uses leading to different effects, curriculum design and implementation related variance of technology adoption and use, problems of the pedagogical approaches of the teachers, and issues that relate to the identification and use of research instruments as well as the interpretation of results generated from their use. In addition, it can also be argued that there are other challenges that are more practical in their nature such as the lack of investment and interest from educational technology companies in measuring the impact of their products on educational outcomes, strategic challenges such as the lack of cross-sector learning and multi-disciplinary collaboration in educational technology research and practice

(between developers, designers, entrepreneurs, educators and academics). To exemplify the first point, for any pharmaceutical product to be available in the market, it has to go through a very strict evaluation process and impact trials. Yet, educational technology products can go straight into the market as soon as they are developed without any evaluation of their impact on their expected educational outcomes, or identification of their potential detrimental effects. To make it clear, we do not have any purpose of comparing medicine as a discipline to education, however, the point we are making is that, at the moment, there is a great uncertainty around the impact of educational technologies on their expected outcomes and this should be decreased to a certain extent through appropriate impact evaluations. This approach would be unacceptable, unethical and illegal in various disciplines including medicine, agriculture, engineering, or architecture. However, yet it appears to be acceptable in educational technology contexts. The demand regarding the decreased uncertainty around the impact of educational technologies would also encourage educational technology producers to be more interested in and invest into impact evaluations. Regarding the second point, efforts to ensure that promising research has at least some chance of being "translated" into useful products are explicit in many other disciplines (see for instance <https://www.elsevier.com/connect>). On the other hand, attempts to create a shared understanding between the key stakeholders in educational technology, including technology developers, academia, and educators, is extremely rare (see for instance <https://educate.london>).

It clear that generating evidence about the impact of educational technologies is a challenging enterprise. Furthermore, evaluating the impact of emerging technologies, which is the particular focus of this chapter, is even more challenging. In this chapter, we discuss two main reasons for this argument. Firstly, compared to more mature and established educational technologies, it is more challenging to explicitly diagnose and identify the actual educational 'problem' specific to the contexts of emerging educational technologies, which also makes outcome measure(s) and their relationship to the expected impact hard to identify. However, transparency of interventions in terms of outcome measures, their definition, and contextual factors are fundamental for the emergence of evidence in educational research (Cukurova, Luckin, & Baines, 2017). In some cases, the solutions created through the emerging educational technology do not relate to any particular educational outcome which would be the key outcome measure of the impact evaluation of the emerging educational technology created. This leads to the result that neither the emerging educational technology interventions created nor the evidence of their impact on certain outcomes has much value for practitioners and policy-makers. Evidence presented in previous reviews (see, for instance, Cox, & Abbott (2004)) shows that likely effects of technology interventions on learning are often present *only when* outcome measures are closely linked to the actual expected learning outcomes of the technology.

Secondly, there has been a tendency in recent educational research to move towards experimental research designs that aim to generate evidence to inform teaching practice while undervaluing other types of research approaches (see for

instance Slavin, 2017). This movement towards positivist research methodologies in education has led to heated debate over the last few years about the way in which evidence in education is perceived, and the relationship between research and practice in educational research (de Bruin, 2015), with much discussion focusing on how to generate a positive relationship between educational research and teaching knowledge and practice (Pampaka, Williams, & Homer, 2016). Here, we argue that due to the ever-changing dynamic nature of emerging educational technologies, the traditional methods of more positivist impact evaluations, such as randomized control trials (RCTs), should *not* be considered as *the only* solution. Such research methods require mature practices, long-term implementations, and big sample sizes which are all problematic to satisfy for emerging educational technologies. In measurements of emerging technology impact, it might be a more appropriate approach to collect evidence in an iterative and adaptive process, disseminating findings as they evolve, and contribute to the maturation of both technology and its implementation in particular educational contexts. After such maturation of technology and its implementation has been reached, the value of more positivist approaches becomes more apparent in the impact evaluations. Traditional positivist impact evaluations cost a significant amount of resources and they require long treatment periods to be meaningful. They do have a great value for more established interventions, which are refined with pilots and early research that use post-positivist approaches with both qualitative and quantitative data. However, for emerging technologies, their value is open to criticisms. Here, we will argue that small-scale rapid cycle evaluation type of investigations might be more appropriate for most of the emerging technologies until they reach a certain level of maturity.

2.1 Transparency of the Intervention and Identification of the Outcome Measure(s)

As we mentioned above, the diagnosis and identification of the actual educational 'problems' specific to the contexts of emerging educational technologies is a challenging but a necessary step in impact evaluations. This challenge might lead to the omission of key contextual factors and clear outcome measures for the impact evaluations of emerging technologies. The omission of contextual factors is hard to understand when one considers that a wide range of research has illustrated that it is impossible to understand how people work or learn without also taking into account the people and artefacts that make up their context (Nardi, 1996). Although the term context is probably the term that is used most frequently within educational research papers to index the circumstances in which learning takes place (Cole, 1996), it is still a complex concept with its various definitions and interpretations by educational researchers. It is very difficult to 'pin-down' in a way that enables context to be used as the basis for informing the practice of teaching and learning. However, the context has been subjected to two principal conceptions: the first conceptualization is 'that which surrounds', which is open to the criticism that context is portrayed as a container rather than part of the same situation. The second conceptualization requires that we interpret mind in a relational way: 'as distributed in the artefacts which are woven together and which weave together individual human actions in concert with and as part of the permeable, changing,

events of life' (Cole et al., 1996). In this chapter, our interpretation and use of the word context are similar to the latter conceptualization of context. In this sense, context is a reflection of the interactions that the learners have experienced with multiple people, artefacts, and environments. This interpretation of context includes a range of factors such as the learning tasks and the learning interactions of social constructivist learning processes, as well as its relatively simplistic interpretation as being that which surrounds learning and learners. It is sadly true that the treatment of context in educational studies, particularly when it is considered in this broader sense described above, is "under-examined, under-theorized and under-developed" (Gulson & Symes, 2007). Context is often so poorly reported within the educational technologies research literature, that a sustainable and systematic manner in measuring emerging technology impact in primary and secondary schools evades the research community. The omission of contextual information from much educational technology research devalues the impact of this research both on the practice of educators and the practice of policy-makers (Cukurova, Luckin, & Baines, 2017).

In the past, due to limitations of technology and its use in education, impact evaluations were often completed through straightforward questions such as "what are the effects of this physics teaching software on students' understanding of the concept of gravity?" or "Is there an impact of students' understanding of chemical bonding after engaging with this simulation?" etc. The evaluations were also undertaken with traditional methods such as pre- and post-test evaluations (Cox, 2008). However, in the case of emerging technologies, the aim is often to transform students' experience of traditional education. Therefore, what exactly this experience involves, including the expected outcome measures and contextual factors, should be clearly defined as the initial step of the impact evaluation process. If we take the position that the context is the sum of one's experiences, then the context in which the impact measured should be as transparent as possible in order to be able to clearly identify the outcome measure(s) which will be investigated during the impact evaluations of emerging technologies. The complex design of emerging educational technologies requires much more understanding of human-computer interactions (Cox, 2005) as well as the wider learning context in which they are being implemented.

Emerging technologies vary enormously and multiple researchers have made clear that the design and use of an educational technology plays a big role in its impact on educational outcomes (see for instance Reeves, 2008; Pilkington, 2008). Not all emerging technologies are equal in their potential to afford efficacy. Any kind of impact evaluation in educational technology research, therefore, also requires detailed knowledge of the nature of the evaluated technology, their different representations and the ways in which they may contribute to learning (Pilkington, 2008). There could be various methods to bring this transparency of the emerging technologies and their contexts, next we will present one of them.

2.1.1 A Theory of Change for Emerging Technologies

Change is the essence of emerging educational technologies. This change involves changes in the technology as well as the changes in the experience of learners. According to the theory of change, until a change occurs, a state of equilibrium exists in which any forces that might drive change are equal to the forces that are resisting change. For change to happen, the balance in the equilibrium needs to be upset (Fullan, 1983). This imbalance can be achieved either by strengthening the driving forces behind the change or removing the barriers that resist the change (Fullan, 2001). In the context of emerging technologies, a theory of change can essentially be represented as a diagram that explains how an emerging technology might have an impact on its users. It should mainly outline all design features that an emerging technology has, the ultimate impact that it aims to have on its users and all the potential outcomes that lead or contribute to this ultimate aim. A theory of change diagram is very useful in impact evaluations as it enables one to identify the steps that are necessary to take in order to reach the expected educational outcomes as well as the assumptions taken for granted between those steps. It also ensures that these outcomes may actually lead to emerging technology's overall impact and that the impact is measurable. The transparency that a theory of change brings into the problems that an emerging technology aims to tackle in education is the first step to take for impact evaluations. There are five main steps of creating a theory of change:

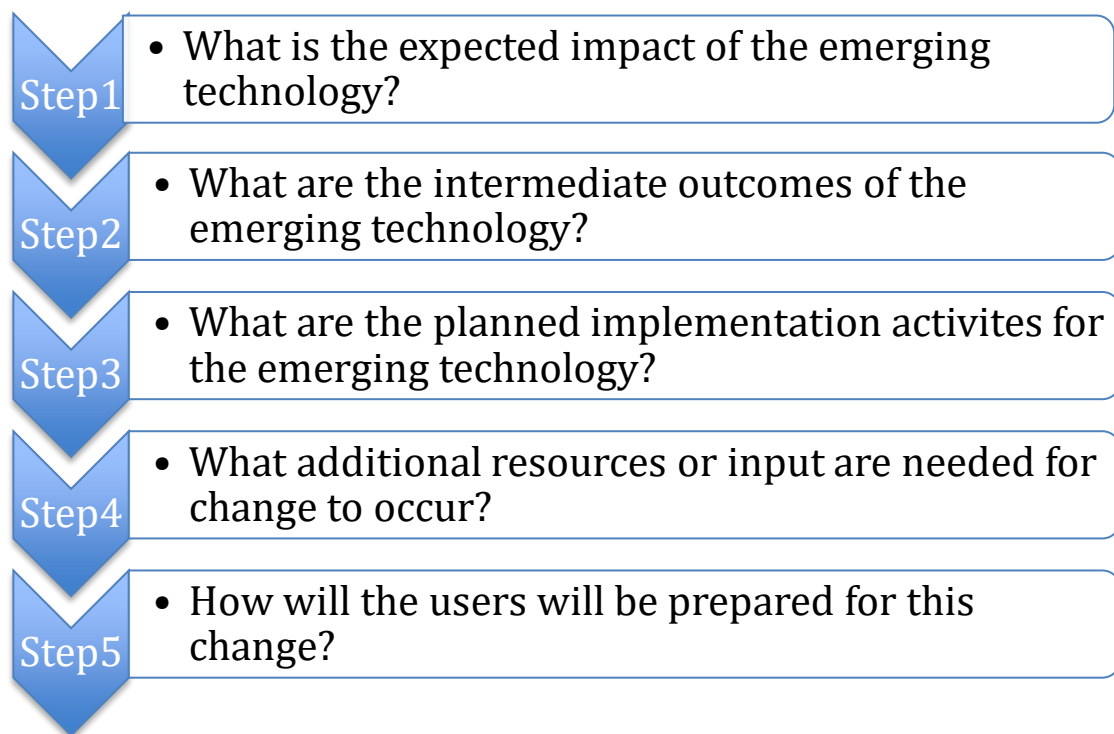


Figure 1: Theory of Change Diagram Steps for Emerging Educational Technologies

Step 1: This step can be considered as the identification of the expected primary impact of the emerging technology. In other words, to what issue in education this emerging technology product might provide a solution to? One of the key challenges here is that impact should be measurable, and ideally, the impact should be the direct impact of the emerging technology that can be observed in realistic timescales. There is no point in creating multiple impact scenarios that are not

specific to the emerging technology, therefore the impact should ideally be one to three bullet points that are specific.

Step 2: These intermediate outcomes lead to the overall impact and can be defined as the expected change/benefit that your users of the emerging technology will experience. These also should be measurable, however, unlike impact, these outcomes should be comprehensive in order to be able to explain all expected changes the users will experience as part of their engagement with the emerging technology.

Step 3: There is great evidence that the activities in which the technology in education is being used moderately the impact on expected learning outcomes (Cox et al., 2004). Activities in which the emerging technologies are being used might range from using what are essentially traditional methods with only small interaction with the technology to more fundamental changes in the teaching and learning practices. For instance, a natural language processing artificial intelligence (AI) tool (such as Amazon Alexa) can be used by a teacher as part of her teaching practice, essentially as a handy dictionary to define unknown words that might emerge during her teaching activities. However, it could also be used as part of students' group activities in which students talk with the AI agents in order to improve their pronunciation, motivated to be understood by the natural language processors of the AI agents (Underwood, 2017). Clearly, the impact evaluations of these two interventions with the same technology would be significantly different. Therefore, in impact evaluations, clear explanation of the implementation activities of the emerging technologies is very important. Activities may comprise everything that the emerging technology does for its users and everything that users do when they are using it as part of the intervention. Activities should only include things that influence the users directly, not operational tasks that are done in advance of the teaching and learning activities or that might impact the user in an indirect manner. As part of a theory of change, activities and their outputs need to be linked to intermediate outcomes identified in step 2. It is useful to draw arrows indicating this potential causality on a theory of change.

Step 4: There are often other resources needed to use the emerging technologies – other than the technology itself. These could include access to the internet or other software or hardware requirements. Likewise, other inputs, such as particular skills or training, might need to be present for successful implementation of the emerging technologies in education. It is important to bring transparency to these resources as part of a theory of change.

Step 5: In addition to making sure your users have the resources and skills necessary, there might be many other challenges with the process of change, especially in education. It is important to consider these when planning for impact evaluations, as they would affect the size of the impact. This step is particularly useful for the interpretation of impact measurements as well as the process of it. A few issues to consider are presented below;

- a) **Readiness for change:** Change in education is a slow process, compounded by the fact that many educational practices haven't changed in decades. Ensuring that the users of the emerging technology are ready for the change, in terms of understanding and accepting the need for the change, is vital for its impact.

- b) **Urgency or need for change:** When processes and practices are so intertwined, it is often difficult for participants to understand why change is necessary. Clear explanations and data that demonstrate why the change is positive and necessary can help convince stakeholders of the need for change.
- c) **Learning anxiety:** Participants need to feel that the conditions are safe for a change. One way to combat this anxiety is to include practice phases as part of the implementation so that participants can test the change and iron out mistakes.
- d) **Implementation dip:** There is good evidence that confidence in using technologies is vital for impact on educational outcomes (Condie, & Munro, 2007). When people realise that they are required to learn new skills and that they aren't as confident in the new practices as they were in the old, an implementation dip occurs. This is a decrease in both performance and confidence during the change process. The implementation dip is to be expected, especially in education and requires coaching and other support to help people through this transition.

2.2 A pragmatic approach to impact evaluations

Impact evaluations of any emerging technology require the generation of evidence regarding any effects arising from an educational intervention that involves that emerging technology. However, views about what constitutes “evidence” may vary considerably among and between stakeholders. For instance, although most educational technology developers would present quotes from their users as evidence of the impact of their product or service, such anecdotal instances would not impress most academics. This variance brings more confusion into the already complex endeavor of educational technology practice, and decreases the likelihood of employing evidence-informed practices in educational environments. When the discussion is on measuring impact, evidence is often categorized in four groups as elaborated below (Hoeken, 2001). It is important to note here that the type of evidence does not necessarily reflect the quality of the evidence and different types of evidence have different advantages and disadvantages (Marshall, & Cox, 2008). Different types of evidences’ quality is judged with different criteria and different types of evidences are more appropriate for different research purposes and contexts. Exploration of the quality criteria for different types of evidence is outside the scope of this chapter (for further information please see O’leary, 2004). Nevertheless, the point that requires to be made here is that, each type of evidence should be judged with appropriate evidence quality criteria and a type of evidence’s appropriateness should be considered for specific questions and within particular research contexts.

2.2.1 Types of Evidence

There are four main categories of evidence: anecdotal, descriptive, correlational, and causal evidence.

(1) Anecdotal evidence is evidence from personal statements or claims based on one or more people's personal experiences. When compared to other types of evidence, anecdotal evidence is generally regarded as limited in value due to a number of potential weaknesses regarding the evidence quality measurement values such as subjectivity. Therefore, it is hard to argue about the impact of emerging technologies using only this type of evidence. However, anecdotal evidence can be considered within the scope of the scientific method as long as it satisfies some of the quality of evidence criteria (auditable and transparent), empirical and verifiable. Often the real value of anecdotal evidence is in its potential to indicate the potential context in which an emerging educational technology can be effective. For instance, in a recent case study, Perry (2015) investigated the potential impact of a new augmented reality-based mobile learning tool which was created for first-year University French students in order to bridge the gap between gaming and education through quest-based learning. As part of the study, volunteers were sought from first-year University level French-language classes and a total of 11 students participated in the study. Based mainly on student excerpts, the study concludes that game-based mechanics can be positive motivators for learners. While this study provides an interesting piece of evidence that could be used as a supplement to a larger body of evidence, it is too limited in scope and sample to stand on its own. However, it addresses several benefits and limitations of the emerging educational technology investigated via a case study, and it offers theoretical and practical implications that will shape the future more targeted studies. For instance, the evidence generated from the case study is used to shape further analysis of student interactions with each other and their interactions with the tool in order to provide a comprehensive analysis of the emerging technology's use and its impact on student motivation.

(2) Descriptive evidence stems from the summary of characterizing individuals and groups who use a specific emerging technology and depicting events, processes, trends, or relationships that emerge from users' interaction with an educational technology resource. This type of evidence can provide more insight into effectiveness; however, the main criticism of this type of evidence is that it doesn't control the other variables that could potentially have an impact on the results generated (Marshall, & Cox, 2008). Educational technology research and commercial documents are replete with such evidence. (see for instance, Beraza, Pina, & Demo, (2010)).

(3) Correlational evidence is the identification of the relationship between a condition or initiative and a specific outcome. When the results show correlations these might be due to the intervention or another causal factor. The correlation results, therefore, cannot answer why questions, yet they are still very useful for evidence-informed predictions.

For instance, Bakker, Heuvel-panhuizen & Robitzch (2015)'s longitudinal study, which is conducted over two years, includes a much larger sample size than the aforementioned case study (719 pupils from 35 primary schools) in order to examine the effectiveness of computer "mini-games" in primary education. Such game

applications can barely be considered as an emerging technology, however, due to lack of longitudinal research on emerging technologies, potentially due to aforementioned dynamic nature of these technologies, we will use this example here. This work is labelled as a controlled longitudinal study. The three groups researched included pupils who used the computer games at home, with no in-school discussion, pupils who used the games at home but discussed them in school, and pupils for whom the games were played in school as a part of the mathematics curriculum. The study found that the games were most effective in improving the multiplicative skills of students who used them at home with follow-up at school.

In designing this study, researchers considered several meta-analyses of the use of technology in mathematics education and included a control group so that the variation between the control group and experimental groups could be attributed to the influence of the technology intervention (in this case, the mathematics mini-games). (Bekker, Heuvel-Panhuizen, & Robitzch, 2015). Due to its relatively large sample and somewhat controlled nature of the study, it provides a nice example of correlational evidence and its value in terms of providing a relatively high level of confidence in the results generated. Even so, the researchers point out the fact that this study cannot be generalized beyond the sample and experimental conditions used here. One of the reasons for this is that their sample was not representative of Dutch schools as a whole, due to the dropout in participation they experienced from their original. Also, the students were not distributed randomly into groups which brings doubts about the equivalence of students in control and intervention groups in terms of potential confounding factors such as their motivation and confidence.

(4) Causal evidence: Strong causal analysis can only be achieved by ensuring that the only difference between the group that receives an educational technology intervention and the comparison group is the intervention itself. Many guidelines for evidence-based practice indicate that the best-quality evidence comes from randomized controlled trials (RCTs). RCTs assign large numbers of subjects into either a control or a treatment group by chance. The treatment group then receives the educational technology intervention being tested and the control group does not. An RCT is one of the strongest forms of design for an impact evaluation in general, because it establishes whether a specific intervention caused an outcome or not.

Evidence of the causality of an intervention is relatively rare in educational technology research, particularly that involving emerging technologies. In order to exemplify this type of evidence, we will have a look at the research conducted by Miller & Robertson (2011). The main objective of their research was to create a large-scale study that examined the impact of the use of computer games in schools on pupils' mental computation skills and mathematics self-concept. In this study, they collected data from 634 primary pupils across 32 schools in Scotland. Pupils were randomly assigned at a school level to experimental or control conditions. In the experimental schools, pupils spent 20 minutes per day for nine weeks on a mathematics brain training game. Control schools simply progressed with normal activities in mathematics classes.

The Miller & Robertson study found significant gains in both groups from the pre- to post-test periods of the research, but the gains of the experimental group were 50% greater than the control group in terms of accuracy and twice that of the control group in speed. In contrast to the previously cited longitudinal study, the authors of this research asserted that their findings can be generalized. However, even in this case of strong evidence, it is worth to mention that the findings would only apply to a restricted part of the curriculum that is investigated in this research. Furthermore, Stern (2015) argues one can draw conclusions from a causal inference based on an RCT that a specific intervention led to an increase in expected learning outcomes. However, when it becomes evident that the same intervention does not always lead to the same learning outcomes in all places, people start to ask 'why?' Another important issue in emerging technologies research is that it is hard to recruit large numbers of participants that are needed to achieve the statistical power required to iron-out various differences between individuals. Regardless, RCTs are an essential part of impact evaluations due to their power to generate reliable causal evidence.

(5) Meta-level evidence: The confidence in evidence studies increases through their meta-level analysis. For instance, Li and Ma (2010) reviewed studies on the impact of technology in general, and on the mathematics learning in primary and secondary classrooms. From existing literature, the authors identified 46 studies encompassing 36,793 learners. Nearly all studies were well-controlled, using a random assignment of students to either experimental or control conditions. In their analyses of pre-existing research, the authors found statistically significant positive effects of the use of technology on students' achievement in mathematics, as well as several other findings that numerous reviewed studies have in common. However, Bakker et al. (2015) note one of the challenges with this meta-analysis is that it only looks at "computer technology" in general, rather than a specific intervention, and thus is difficult to apply the results of such a metalevel evidence into specific emerging technologies. In addition, as also argued by the authors, "It is impossible for any meta-analysis to evaluate the design quality of the programs used in primary studies" (Li & Ma, 2010; p. 235). Due to the challenges of identifying individual emerging technologies from review studies and the lack of details about the design quality of the technologies in primary studies, values of such metalevel investigations in practice is limited. These issues emphasise the value of a clear theory of change in impact evaluations (see section 2.1.1).

It is also worth to make it clear that meta-level evidence does not have to come from quantitative studies (see for detailed discussion Noblit, & Hare, 1988). For instance, Tondeur et al. (2012) reviewed 19 qualitative studies that focused on strategies to prepare pre-service teachers to integrate technology into their lessons. In their synthesis, they explain key themes explicitly related to the preparation of pre-service teachers (e.g., using teacher educators as role models, learning technology by design, scaffolding authentic technology experiences), and conditions necessary at the institutional level (e.g., technology planning and leadership, co-operation within and between institutions, training staff). The meta-level evidence generated then used to create an overarching model that explains the relationship between these factors. More extensive meta-studies are also discussed in Chapter 2 of this section.

It is increasingly clear then, that different types of evidence in impact evaluations of emerging technologies have different advantages and disadvantages. Therefore, relying on one study or one type of evidence is unlikely to provide enough reliable evidence to judge the impact of an educational technology. Unfortunately, for most traditional experimental methods, once the generalization has been made, the assumption, particularly among practitioners, is that the result will occur regardless of the means of getting there, both in terms of the context and the process. This results in issues that even though well-established evidence regarding the impact of emerging technologies on various learning outcomes exists, it is hard to argue that the evidence would be strong under every contextual factor or even so, its strength would be the same (Cukurova, Luckin, & Baines, 2017). In addition, as argued by Kelly, Baek, Lesh, and Bannan-Ritland (2008) emerging technology interventions should respond to new knowledge emerging from recent research studies. This is not common in more traditional educational research and might even be considered as unproductive. However, given the pace of change in emerging technologies, it is not always easy for researchers in the field to get satisfying support from the findings of the effectiveness of previous interventions investigated. As one potential solution to this, Ann Brown (1992) discusses strategies to transform schools into communities of learning and interpretation where students are provided opportunities to take charge of their own learning and calls for new and complex methodologies to capture the systemic nature of learning, teaching, and assessment in their own contexts. Such approaches to empowering other stakeholders of educational technologies rather than relying only on researchers in impact evaluations might potentially be valuable for measuring emerging technologies' impact in primary and secondary education (see for instance <https://educate.london>). Similarly, more recently, Design-Based Implementation Research (DBIR) aims to organise research and development that addresses the challenge of creating effective, scalable, and sustainable policies and programs in education. The approach argues that continuous close connection between researchers, teachers and their local education districts is fundamental for effective implementation of educational interventions (Coburn, Russell, Kaufman, Stein, 2012).

Multiple sources of evidence are needed in order to strengthen the argument that a particular technology intervention will be successful under a variety of conditions. Both quantitative and qualitative sources of evidence are valuable for the statistical power of their large sample sizes (correlational and causal evidence) and the explanatory power of more in-depth questioning (anecdotal and descriptive evidence). It would be premature if decisions were taken about whether or not to implement an intervention based on one type of evidence only. A more holistic approach is needed in order to reach a complete picture regarding the impact of emerging technologies in education. All evidence types may shed light on why an intervention of an emerging technology succeeded or not, and in what circumstances. Rather than arguing about the overall superiority of one particular type of evidence or research approach over others, perhaps a more important questions to ask are what type of evidence is the most appropriate for a particular emerging technology innovation? and how can we design and implement interventions that might help us generate this type of evidence? Such questions are

particularly important in domains with little prior research (Cobb et al., 2003) which is, by definition, very often the case for emerging educational technologies.

2.2.2 Innovation stages of emerging technologies

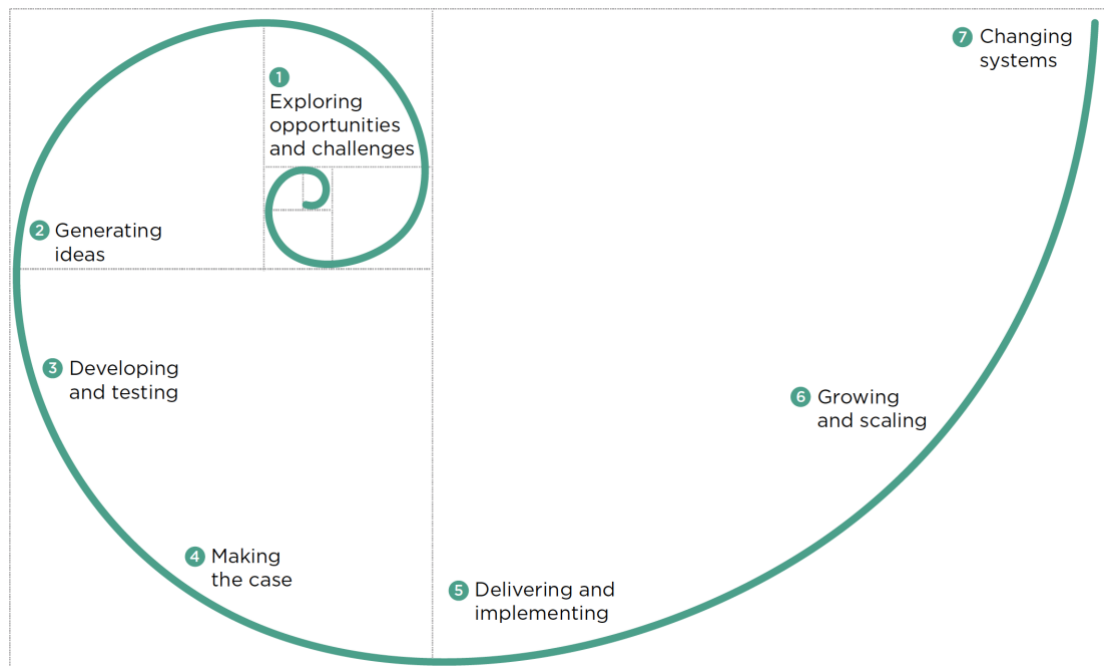


Figure 2: Innovation cycle as presented in *Using Research Evidence: A Practical Guide* (2016)

The spiral above was developed by Nesta (2016) to capture the different stages of the innovation process and it can be used to identify different innovation stages of emerging technologies. As argued by Breton et al., (2016) different stages of innovation would require different types of evidence. For instance, initial stages of exploring opportunities and challenges, as well as generating ideas, it would be beneficial to focus on literature reviews and design principles, identifying what has worked or failed in the past in different contexts and using this evidence in the design decisions made for the emerging technologies. These design principles and lessons can help both developers and users of emerging educational technologies follow strategies that are more likely to have an impact. During the developing and testing stage, rapid cycle evaluations that would generate anecdotal and descriptive evidence would be beneficial, whereas at making the case stage it would be beneficial to undertake impact evaluations that would generate some correlational evidence. Once an emerging technology reaches certain level of maturation through these stages, during the delivery and implementation stage, it would require causal evidence that would show causal impact. On the other hand, growing, scaling and spreading stage would require bigger scale experimental evaluations. System-level change can only be provided through multiple big scale evaluations from various contexts and clear implementation manuals that would ensure impact in multiple places. It is interesting to note here, that by the time an emerging technology reaches to system changing level, it would have reached a certain level of maturation so much so that its emerging nature would be questioned.

The approach put forward by Nesta considers evidence in a holistic manner and recognises the value of different types of evidence at different stages of emerging technology innovation. It contrasts with the more traditional approach of considering evidence types in a hierarchical manner, in which causal research evidence is considered as 'gold standard' and other types of evidence are undervalued. We argue for the synergy of evidence types and research methodologies to generate different types of evidence during impact measures of emerging technologies. This position is based on the view that kite-marking a certain technology as 'effective' based on 'gold standard' causal evidence, and encouraging its scaling might be a futile approach for emerging technologies. As mentioned in the introduction to his chapter, emerging technologies are constantly evolving and being implemented in different contexts with different populations. Therefore the value of previous experimental evaluations for an emerging technology is limited. In addition to this, meaningful big scale positivist evaluations of emerging technologies are expensive and they take a long time to be completed. There are various research methodologies that can produce *valuable indicators of the potential impact* and they should be encouraged before encouraging researchers, emerging technology companies and users for such big scale evaluations.

3. Conclusions

Emerging technologies have a vitally important role to play in any change management in education. Emerging technologies can also disrupt and bring about unexpected change, the consequences of which must then be managed. Either way, the evaluation of these emerging technologies is a key part of the way in which their impacts are effectively integrated into learning and teaching settings to bring the best benefit to learners and teachers. Artificial Intelligence technologies for use in education are a good example of an emerging technology that will inevitably bring change. Educators, learners, parents and policymakers alike, therefore, need reliable methodologies for evaluating the effectiveness of such emerging technologies.

Discussions about the impact of emerging educational technology should focus on the suitability of the proposed evaluation methods, and it should focus on the types of evidence rather than on the hierarchy of these methods and evidence types. Different impact evaluations have different purposes. It is therefore important to recognise that any research design and methods can only meet the requirements of a subset of the possible evaluation purposes. We, therefore, argue for an approach that focusses on the identification of the specific research questions that are appropriate for the educational issues that are relevant to the context of the intervention to be evaluated. Finding the right research method to generate the right type of evidence is of crucial importance. Impact evaluations of emerging technologies should be considered as a process, rather than as a case if we are to progress to a more productive way forward in evidence-informed educational technology discussions.

In this chapter, we have introduced an approach that can be adopted as one potential process of measuring the impact of emerging technologies in education.

Our approach has two main steps. First one is the creation of a clear theory of change to identify outcome measures and assumptions that are behind the expected impact of the emerging technology intervention. Secondly, the identification of the type of evidence and methods to generate it that are the most appropriate for the current innovation stage of the emerging technology. We argue that the more traditional approach of generating causal evidence through empirical research studies and argue for the 'effectiveness' of an intervention in order to scale it might not be a productive approach for emerging technologies. Rather we suggest that during impact evaluations of emerging technologies all types of evidence both generated from intervention trials, case studies, and existing literature should be leveraged at different innovation stages of emerging technologies.

References:

- Ainley, J., Enger, L., & Searle, D. (2008). Students in a digital age: Implications of ICT for teaching and learning. In J.Voogt & G.Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 63-80). New York: Springer.
- Bakker, M., Heuvel - Panhuizen, M., & Robitzsch, A. (2015). Longitudinal data on the effectiveness of mathematics mini - games in primary education. *British Journal of Educational Technology*, 46(5), 999-1004.
- Beraza, I., Pina, A., & Demo, B. (2010, November). Soft & Hard ideas to improve interaction with robots for Kids & Teachers. In workshop Proceedings of Conference on simulation, modelling, and programming for autonomous robots (pp. 549-557).
- Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational theory*, 57(1), 1-22.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences*, 2(2), 141-178.
- Buck, S., & McGee, J. (2015). *Why government needs more randomized controlled trials: refuting the Myths*. Laura and John Arnold Foundation, Houston, TX, arnoldfoundation.org.
- Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational researcher*, 32(1), 9-13.
- Coburn, C. E., Russell, J. L., Kaufman, J. H., Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education*, 119 (1), 137-182.
- Cole, K. C., VanTilburg, D., Burch-Vernon, A., & Riccio, D. C. (1996). The importance of context in the US preexposure effect in CTA: Novel versus latently inhibited contextual stimuli. *Learning and Motivation*, 27(4), 362-374.
- Condie, R., & Munro, B. (2007). *The impact of ICT in schools: Landscape review*. UCL Institute of Education.
- Cox, M. J. (2005). Educational conflict: The problems in institutionalizing new technologies in education. In G. Kouzelis, M. Pournari, M. Stoepler, & V. Tselfes (Eds.), *Knowledge in the new technology*. (pp. 139-165). Frankfurt: Peter Lang.
- Cox, M. J. (2008). Researching IT in education. In J.Voogt & G.Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 965-981). Springer US.
- Cox, M.J. & Marshall, G. (2007) Effects of ICT: do we know what we should know? *Education and Information Technologies*. 12: 59-70

- Cox, M.J., Webb, M., Abbott, C., Blakeley, B., Beauchamp, T., & Rhodes, V. (2003). *ICT and pedagogy: A review of the research literature*. ICT in schools research and evaluation series, 18.
- Cox, M. J., & Abbott, C. (Eds.) (2004), ICT and attainment – a review of the research literature. Full Report. Becta Coventry/DfES London January 2004
<http://www.becta.org.uk/research/index.cfm>
- Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education*, 110, 51-63.
- Cress, U., Moskaliuk, J., & Jeong, H. (Eds.). (2016). *Mass collaboration and education* (Vol. 16). Springer.
- Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education*, 116, 93-109.
- Cukurova, M., Luckin, R., & Baines, E. (2017). The significance of context for the emergence and implementation of research evidence: the case of collaborative problem-solving. *Oxford Review of Education*, 1-16.
- De Bruin, C. L. (2015). Conceptualizing effectiveness in disability research. *International Journal of Research & Method in Education*. doi:10.1080/1743727X.2015.1033391
- Department for Education. (2014). *School funding: Research priorities and questions*. Crown copyright.
- Department for International Development (2017). Public Bulletin Board 7749: Education Technology to Improve Learning for All. Retrieved from
<https://supplierportal.dfid.gov.uk/selfservice/pages/public/viewPublicNotice.cmd?bm90aWNISWQ9Njc3MjA%3D>
- Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J.Voogt & G.Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 735-745). Springer New York.
- Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.
- Fullan, M. (1983). Evaluating program implementation: What can be learned from Follow Through. *Curriculum Inquiry*, 13(2), 215-227.
- Greany, T., & Maxwell, B. (2017). Evidence-informed innovation in schools: aligning collaborative research and development with high quality professional learning for teachers. *International Journal of Innovation in Education*, 4(2-3), 147-170.
- Gulson, K. N., & Symes, C. (2007). *Spatial theories of education: Policy and geography matters*: Routledge.
- Katterfeldt, E., Cukurova, M., Spikol, D., Cuartielles, D. (2018). Physical Computing with Plug-And-Play Toolkits: Key Recommendations for Collaborative Learning Implementations. *International Journal of Child-Computer Interaction*, 1-23.
- Kelly, A. E., Baek, J. Y., Lesh, R. A., & Bannan-Ritland, B. (2008). Enabling innovations in education and systematizing their impact. In A.E. Kelly, R.A. Lesh, J.Y. Baek (Eds) *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching*, 3-18.
- Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, 22(3), 215-243.
- Marshall, G., & Cox, M. J. (2008). Research methods: Their design, applicability and reliability. In J.Voogt & G.Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 983-1002). Springer, Boston, MA.
- McDougall A. Murnane, J., Jones, A. & Reynolds, N. (Eds.), *Researching IT in Education: Theory, Practice and Future Directions*. Routledge: Taylor and Francis Group: Abingdon. England

- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29-40.
- Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology*, 42(5), 850-864.
- Nardi, B. A. (1996). Studying context: A comparison of activity theory, situated action models, and distributed cognition. *Context and consciousness: Activity theory and human-computer interaction*, 69-102.
- NESTA (2016). Using research evidence for success: A practical guide. Retrieved on from https://www.nesta.org.uk/sites/default/files/using_research_evidence_for_success_-_a_practice_guide.pdf
- Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, California: Sage.
- O'leary, Z. (2004). *The essential guide to doing research*. Sage.
- Pampaka, M., Williams, J., & Homer, M. (2016). Is the educational 'what works' agenda working? Critical methodological developments. *International Journal of Research and Method in Education*, 39, 231–236.
- Perry, B. (2015). Gamifying French Language Learning: a case study examining a quest-based, augmented reality mobile learning-tool. *Procedia-Social and Behavioral Sciences*, 174, 2308-2315.
- Petty, G. (2009). *Evidence-based teaching: A practical approach*: Nelson Thornes.
- Pilkington, R. (2008). Measuring the impact of IT on students' learning. In J. Voogt, & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education*. Berlin Heidelberg New York: Springer.
- Reeves, T. C. (2008). Evaluation of the design and development of IT tools in education. In J. Voogt, & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education*. Berlin Heidelberg New York: Springer.
- Slavin, R. E. (2017). Evidence-based reform in education. *Journal of Education for Students Placed at Risk (JESPAR)*, 22(3), 178-184.
- Stern, E. (2015). *Impact evaluation: A guide for commissioners and manager*. Big Lottery Fund and the Department for International Development, UK.
- Tondeur, J., van Braak, J. Guoyuan, S. Voogt, J., & Fisser, P. & Ottenbreitt-Leftwich, AS. (2012). Preparing student teachers to integrate ICT in classroom practice: a synthesis of qualitative evidence. *Computers & Education*. 59,1, 134-144. doi.org/10.1016/j.compedu.2011.10.009
- Underwood, J. (2017). Exploring AI language assistants with primary EFL students. *CALL in a climate of change: adapting to turbulent global conditions—EUROCALL 2017*, 317.
- Voogt, J., & Knezek, G. (Eds.) (2008) *International Handbook of Information Technology in Primary and Secondary Education*. Berlin, Heidelberg, New York. Springer