Using evidence-informed logic models to bridge methods in educational evaluation

COLDWELL, Michael <http://orcid.org/0000-0002-7385-3077> and MAXWELL, Bronwen <http://orcid.org/0000-0002-8022-9213>

Using evidence-informed logic models to bridge methods in educational evaluation

Abstract

Designs combining different types of data are increasingly used in educational evaluation, to provide both evidence of impact and an explanation of the processes by which impacts are created. Logic models are visual representations of how an intervention leads via a set of steps from resources and inputs to outputs and then sets of outcomes. Their use has become widespread to underpin evaluations; and they have become of more interest in education as they have been promoted by policy makers and funders including the Education Endowment Foundation (EEF) in England. This paper addresses the question: how can logic models be used to frame and implement educational evaluations using combinations of methods? To do so, the paper draws on theory-based evaluation literature to identify a set of issues to be considered: the role of implementation logic; causal mechanisms; the context of interventions; and the importance of considering and addressing issues around complexity. Using detailed examples from two study designs for EEF evaluations, the paper presents an evidence-informed logic model approach to deal with these issues. The paper concludes by reflecting on the practical and theoretical implications of this approach, laying out a set of key issues to address in future evaluations for which a design framed by an evidence-informed logic model may be appropriate.

Introduction

Designs that combine differing forms of data are increasingly used to structure educational evaluation studies, for a variety of reasons. In particular, using combinations of methods can help improve understanding and enable better interpretation of findings from evaluations with a variety of purposes including impact, pilot and scale-up evaluations, all of which are considered in this paper. The use of logic models (Rogers, 2008) to lay out the steps from inputs to outcomes of programmes has become widespread as a tool for designing evaluations of social policy and interventions and in recent years the use of such models has become of more interest in education as they have been promoted by policy makers and funders including the Education Endowment Foundation (EEF) in England. Yet the use of logic models in educational evaluation, in particular, has not previously been subjected to adequate critical consideration, in the way that such models have been in the wider evaluation field, to support their use to both provide the most robust representation of the intervention being evaluated and to interpret evaluation findings.

This paper addresses the question: how can logic models be used to frame and implement designs that utilise combinations of data in educational evaluation? We consider the potential role of logic models to help in design, analysis and claim-making, drawing on the longstanding theory-based evaluation research literature. In so doing, we present a critique of such models in relation to their utility in explicating implementation paths and causal mechanisms; as well as taking meaningful account of the context and complexity of interventions. From this critique, we discuss an approach to the use of what we term *evidence-informed logic models* that aims to overcome the limitations discussed. Evidence-informed logic models provide a framework for bridging methods as well as bridging differing evaluation traditions - impact evaluation and 'theory-based' evaluation. We illustrate this approach by presenting how we have used evidence-based logic models in two recent EEF evaluations. The paper concludes by reflecting on the practical and theoretical implications of the approaches used in these studies, presenting a new logic model frame that addresses some of the weaknesses in the studies and drawing outa set of key issues to address in future evaluations for which a methods design framed by an evidence-informed logic model might be appropriate.

Terminology

We use the term intervention (rather than program/me often used in the wider evaluation field) to describe the sets of activities and processes that are the subject of evaluation, as this is the usual language used in education. However, when we are referring directly to the work of other researchers we use their preferred terminology.

We use the term theory-based evaluation to describe the set of evaluation approaches that seek to understand the causal logic underlying an intervention, unless directly citing the work of specific researchers in which case we use their terminology. This is particularly important here as researchers in the evaluation field tend to create a range of new terminology for their own variations on the theme ('programme theory evaluation' preferred by Rogers (2008); 'theory-based evaluation' by Weiss (1997); 'realist evaluation' by Pawson and Tilley (1997); and so on - "theory-driven, theory-orientated, theory-anchored, theory-of-change" and more are also identified by Rogers and Weiss (2007, p.63).

We define other key terms as the paper develops.

The development and uses of logic models to underpin evaluation

In this section, we discuss the development of the logic model as a type of causal model representing how an intervention can lead to outcomes. Rogers et al (2000) trace the use of such causal intervention models to guide evaluation back to the 1960s and the work of Sucham (1967) and others, but the key work took place from the 1970s onwards with the development of a group of theory-based evaluation approaches, especially the seminal work of Weiss. Theory-based evaluation covers the set of evaluation frameworks each of which "involves some attempt to 'unpack' the black box so that the inner components or logic of a program can be inspected" (Astbury and Leeuw, 2010 p.364). By the late 2000s, Rogers and Weiss (2007, p.64) noted how things had moved on: they identified that funders now often required evaluation and indeed planning proposals "to include a logic model of program theory from the beginning", and this requirement has slowly seeped into education, driven more recently by the work of the Education Endowment Foundation.

Interestingly, theory-based evaluation developed in response to a significant drawback with impact evaluations using experimental approaches: their weakness in providing an explanation of findings described by Weiss (1997, p.502) as "the usual inability of even the most sophisticated experimental evaluations to explain what factors were responsible for the program's success- or failure. Although evaluations based on random assignment to program and control groups give good estimates of program impact, they have little to say about how or why the impacts occurred."

The use of the term "theory" in this tradition is akin to Merton's 'middle-range' theories: those "that lie between the minor but necessary working hypotheses … and the all-inclusive systematic efforts to develop a unified theory" (Merton, 1968, p.39): it is a description of how an intervention leads to change. Weiss (1998, p.57) describes programme theory thus: "the mechanisms that mediate between the delivery (and receipt) of the program and the emergence of the outcomes of interest". Whilst the term causal mechanism is widely used in the evaluation field in this way as a synonym for causal theory, its meaning varies. Lacouture et al (2015) suggest that its usual use in theory-based evaluation, especially the realist evaluation branch (Pawson and Tilley, 1997; Pawson, 2013), indicates that mechanisms describe processes that result in changes in the *minds* - the thinking and behaviours - of individuals or groups of individuals that are subject to the intervention being evaluated. Thus a mechanism is "hidden but real, is an element of reasoning and reactions of agents in regard to the resources available in a given context to bring about changes through the implementation of an intervention, and evolves within an open space-time and social system of relationships." Our perspective broadly accords with this view, however we would include social mechanisms that link to relations between individuals and groups of individuals, drawing on Charles Tilly's work. Tilly (2005) articulates his concept of 'relational mechanism' most clearly in relation to what he describes as 'boundary mechanisms'. These include, for example, boundary change 'precipitation' mechanisms such as 'encounter' when two previously unrelated groups begin interacting and 'imposition' where authorities draw lines between groups (e.g. grouping pupils in schools); and such boundary change 'enactment' mechanisms such as 'site transfer' e.g. ceremonial transition from childhood to adulthood or achievement of qualification and 'inscription' - formally putting in place laws, rules etc (see Tilly, 2005, chapter 9 for an extended discussion). Therefore, in our work, we consider causal mechanisms as enacted in causal evaluation models as both individual and social in nature.

Within the theory-based evaluation field, various models are proposed to lay out the workings of these causal mechanisms; and as Rogers (2008, p.30) suggests, a typical [program] theory-based evaluation approach uses a "causal model linking programme inputs and activities to a chain of intended or observed outcomes, and then uses this model to guide the evaluation". The logic model is described as "the summarized theory of how the intervention works (usually in diagrammatic form)" that underlies the evaluation. Although largely explicated in the evaluation field, they can be equally useful in the planning arena. Indeed, Rogers et al (2000) define 'program theory evaluation' as having two elements: "an explicit theory or model of how the program causes the intended or observed outcomes and an evaluation that is at least partly guided by this model" (Rogers et al, 2000 p.5). This is helpful as it distinguishes the focus of this section - literature on the logic model as an example of a causal model of how an intervention can lead to outcomes - from this arises how it can be used to guide methods, which we turn to in relation how such models can bridge methods in a later section.

The interest in logic models in the education field intersects with this focus on theorisation in the wider evaluation field, and can be related to two main trends in social research and evaluation that have led to the current positon of widespread use of such models.

Firstly, the recent policy and methodological turn in education to quantitative and trial-based methodologies has led to the development of designs that require explicit causal sequencing of intervention and outcomes. It is instructive to consider changes in EEF's approach over time here. Early EEF trials did not use such models. However, they are now explicitly required by EEF:

> *"Evaluators should develop an intervention logic model or theory of change in partnership with the delivery team to inform the evaluation [...] It is important to know not just if an intervention 'works' in terms of producing desired outcomes, but also if it works in the manner theorised*." *(*Humphrey et al, 2015 p.9).

Secondly, the move towards school-led developments in England, and more broadly, has supported the use of project planning tools. For example, school-led collaborative "Research and Development" projects are often supported by systematic framing of the project and its expected outcomes (Greany and Maxwell, 2017). Since this shift is taking place in the context of the English "school-led self-improving system" (Hargreaves, 2012), seeing knowledge production as potentially part of the role of schools, logic models that draw on prior research evidence are increasingly common.

Typically, in both education and wider evaluation fields, simple logic models are used: single path models between inputs/activities, outputs, and intermediate and longer term outcomes/impact. These are the most widely used of the group of such models, popularised by, amongst others, the Kellogg Foundation and the University of Wisconsin, which provides a useful description of this group of models:

> *"Logic Models are a popular tool that can be used to help conceptualize your change effort. It does this by inviting the author(s) to articulate their understanding of the current situation, the changes they hope to bring about through their program effort, with and/for whom, the activities planned to contribute toward this change, the resources needed to put*

into the effort, assumptions they are making, and external factors that could influence results." (University of Wisconsin, undated)

The University of Wisconsin's simplest version (Figure 1) - developed in evaluations across a range of public policy fields, and widely used, since it is freely downloadable, by evaluators and designers - is fairly typical of the visual representation often used in such models.

Figure 1 here

In this version, inputs are human and other resources provided to enact the programme; outputs are the activities undertaken and outcomes are the intended and sometimes unintended results of the evaluation, such as reduction in crime, improved mental health or pupil attainment. External factors (University of Wisconsin, undated) "include the cultural milieu, the climate, economic structure, housing patterns, demographic patterns, political environment, background and experiences of program participants, media influence, changing policies and priorities" and assumptions relate to the theory or change - "the beliefs we have about the program and the people involved and the way we think the program will work." This kind of single path model is also widely presented in the literature (see for example much of the literature reported in the Canadian Journal of Evaluation; Cox, 2000) alongside other complicated and complex models that can be used as we will discuss.

Typically, logic models are developed using a combination of prior evidence and design principles established by the developers via interviews with key stakeholders or group evaluation (Rogers et al, 2000)- although it is possible for them to use other means; for example, Goertzen, Hampton and Jeffery's (2003) use of grounded theory. The rest of this article focuses on their use in evaluation design, however they are used in other forms of research, in particular organising reviews of literature, for example Kunzle et al's (2010) use of an input-process-output model in a review of patient safety, and in the knowledge mobilisation field - for example Langer, Tripney and Gough's (2016) appropriation from the implementation science field of a COM-B model (Michie, van Stralen and West, 2011). This model examines the differing contributions of literature examined to changing capabilities (C), opportunities (O) and/or motivation (M) to influence Behaviour (B) - in this, case relating to use of research evidence in decision-making, to derive "*a basic logic model that sets out how evidence use interventions are assumed to influence decision-makers' consideration of research evidence […] to structure the interventions according to the applied intervention mechanisms [which] allowed us to create a structure that equally applied to the EIDM and broader social science literature.*" (p11)

Such models have numerous advantages for the evaluator. For example, Rogers et al (2000, p.10-12) provide a very helpful set of reasons for their use, as follows: to understand why interventions do or do not work by articulating their causal logic; to help in attributing outcomes to interventions [especially where randomisation is not possible]; and to help improve interventions by articulating the thinking behind the programme's working and testing this. Cox (2000, p.115) adds that - in relation to the implementation logic element - they can "clarify how a programme is structured for all stakeholders, including funders, and can facilitate programme planning", and this is certainly how logic models focussed on implementation paths are used at the time of writing in the early stages of development of EEF RCTs. However, for their use to be most productive, a set of issues needs to be considered which we draw out from the research literature and our own development of such

models in relation to implementation logic, causal theory, context and complexity which we turn to in the next section.

Issues with logic model designs

The differing roles of causal mechanisms and implementation logic

One key argument first introduced by Weiss (1997) is that the framing of an intervention (or program/me using her preferred terminology) requires a focus both on the theoretical basis for the intervention - what she refers to as the 'program theory' - and how it is implemented - the 'implementation theory'. Weiss (1997, p.506) argues that:

> *"Together, implementation theory and program theory represent the full "theory of change"… Both kinds of theory can be useful, but conflating and confounding them has led to muddy thinking and confusion. Recognition of their distinctiveness would allow an evaluator to capitalize on their different strengths."*

The distinction between programme theory and implementation theory is not always clear-cut, as we will go on to explore in relation to the studies we discuss in subsequent sections; nevertheless, the distinction is useful.

Weiss's terminology can be confusing, especially her use of the term programme theory which has different meanings when used by other theory-based evaluators. So we introduce here three key terms that will be used from here on in this article. First, we use the term **'implementation logic'** to describe the logic behind the series of steps laid out in a logic model implementation pathway or set of pathways. Secondly, we use the term **'causal mechanism'** to describe the explanation/s for why these steps are likely to occur and lead to sought-for outcomes. Thirdly, we use the term **'causal process'** to describe the combination of the two: the implementation path, underpinned by the implementation logic, which is anticipated to lead to sought-for outcomes via a particular causal mechanism.

Drawing on this distinction, it is argued by some in the field that logic models, by explicating the causal mechanisms as well as the implementation logic, can help distinguish between theory failure and implementation failure, as researchers including Lipsey (1993) and Bickman (1996) identify. Cox (2000, p116) explains that theory failure occurs when the intervention is implemented as expected; so "*if outcomes are not achieved, then it can be assumed that assumptions about the impacts or consequences of the programme were incorrect*".

In the next subsections, we look in more depth at these issues - the implementation logic and the causal mechanisms underpinning it - in turn. Then move on to consider context and complexity.

Issues relating to implementation logic

Implementation failure, as described above, occurs where the suggested implementation pathways do not play out as expected - where there is low fidelity to the intervention protocol in trial-based designs, for example. Stame (2010) relates implementation failure to a lack of appreciation of the role of the context within which programmes take place since if an intervention occurs in a context that is not conducive to its implementation then it is not likely to play out in line with the expected

implementation logic. We return to this in more depth in a later subsection, but examples of how this can occur that are commonly found in the education field include a lack of support from school leaders or low motivation of participants (Coldwell, in press).

Possibly the most significant issue related to the implementation logic of interventions is that visual logic models are often presented - as in the case of the Wisconsin model shown above - as single pathways or chains. There are several implications of this. Firstly, as identified by Cox (2000) it can be difficult to identify a simple implementation path, especially where interventions involve multiple, inter-related strands. For example, in Cox's study relating to public health, "the relationships between various health promotion activities and behaviour change are complex and ill defined" (Cox, 2000, p.119).

Secondly, as Weiss (1997) first identified, evaluations using logic models tend to focus on the implementation logic rather than including a focus on the causal mechanisms: "Much of what goes under the label of TBE follows the chain of implementation." Rogers and Weiss (2007, p.64) note that this is widespread and many organisations have "institutionalized a version involving five or so boxes arranged linearly: inputs, activities, outputs, outcomes and impacts" using differing terminology and sometimes additional boxes such as "context, assumptions and external factors" - clearly referring to the popular Wisconsin model amongst others, which "fall short" because they do not examine the causal mechanisms involved. As Astbury and Leeuw (2010, 367) note, “mechanisms appear too frequently as unexplained ‘‘causal arrows’’ that seem to flourish so well in the present climate of enthusiasm with visual logic models.”

Thirdly, simple, single path models can overstate (or understate) impacts if used as the sole basis for evaluation, since they do not take into account context and concurrent programmes and as, Rogers (2008, p.34) indicates, they can focus activity on meeting targets detailed in the model rather than "actual goals of the intervention". Bakewell and Garbutt (2005, p.19) go further in relation to international development, suggesting that simple input-process-outcome models that suggest predictability are inappropriate arguing that they are going beyond a methodology to shape an approach to development: they are "being used as a tool to impose a set of development ideas on communities in developing countries. As such it represents an ideology rather than being an objective, technical management tool".

Wholey (2003) suggests, drawing on some of these criticisms, that only a narrow range of interventions are suitable for a "simple logic model - those where goals can be agreed and precisely quantified, where progress towards them can be reliably measured and where both staff activities [e.g. teaching] and the results of those activities can be readily observed" - a useful point to reflect on for educational evaluators, suggesting that more complex, multi-strand presentations are likely to prove more useful.

Issues relating to causal mechanisms

The first point to make in relation to causal mechanisms is that, as Rogers (2008) identifies, logic models may not only fail to lay out the causal mechanisms involved, even if they do so they often fail to consider competing causal mechanisms to determine which is best supported by evidence: or - more likely - which combination. This is especially true for those involving a single linear pathway. For example Pawson and Tilley (1997, p.78-79) lay out eight alternative mechanisms behind how

CCTV can reduce crime in car parks, combinations of which are likely to occur in differing contexts. Stame (2010) suggests that failing to account for rival causal accounts can lead to programme theory failure, linking this to a failure to deal with complexity, an issue we return to at the end of this subsection. And as Rogers and Weiss (2007) identify, even where causal mechanisms are explicitly considered, often these are poorly evidenced or even discredited such as the "knowledge - attitudes - practice model" - still widely used in educational evaluation.

Secondly, from our own work as evaluators, we would point to the need to consider the role of inter-related but independent causal processes. A common example in relation to EEF pilot evaluations and trials is the combination of a professional development programme that is expected to change teachers' practices with a separate process by which practices in the classroom influence pupil outcomes according to a particular theoretical perspective or evidence base. The current Integrating English evaluation, discussed further below, (Sheffield Institute of Education, 2017) provides a good example from an EEF trial: the key intervention is a change in teachers' practices to improve pupils' functional language skills across the curriculum. The teachers' practice change is intended to occur via a series of workshops culminating in a new scheme of work which supports the changed practice expected. Logically, these are distinct processes with differing outcomes (for the first the outcome is on teacher practices; for the second, the outcome is on pupils' learning or attainment). It is quite possible that a professional development programme might lead to changes in teacher practices, but these practices might not lead to sought for outcomes for pupils. Conversely, the professional development programme might not be effective in changing teacher practices; in which case the evaluation can say nothing about whether such practices may have led to pupil change. Most logic model designs treat these as a single change process with a focus on the end point impact on pupils, but we would argue that this is not necessarily the case, an argument we illustrate in relation to the RETAIN evaluation which we will go on to discuss in subsequent sections.

Dealing with context and complexity

Stame (2004, p.63) identifies that a common position of theory-informed evaluation approaches is to "consider programmes in their context." Greene (2005, p.83) notes that in these designs the context (“the site, location, environment, or milieu for a given evaluand”) is used to help explain how and whether the theoretical model used (both the causal theory and implementation logic) is enacted.

In many visual logic model representations, such contextual features are presented as a box at the bottom of, or in a circle around, the model. For example, the Wisconsin model (Figure 1) includes a box containing 'external factors' that are expected to influence the likelihood of the theory/ies underlying the model being enacted. Elsewhere (Coldwell, in press) we present a detailed argument that logic models tend to oversimplify the nature and role of the context within which an intervention takes place, drawing out a set of features of context that should be addressed. We argue there that consideration of context in logic models should take into a set of key features which we summarise below.

Firstly, logic models present the context for programmes as if they are unchanging: a backdrop or setting within which the intervention occurs. Yet, in the education field for example, organisations are constantly changing, often in a purposeful way. Furthermore, interventions can lead to changes in aspects of the context. For example, an intervention may improve capacity to support the

implementation of new projects, a deliberate aim in some cases - see, for example, the recent EEF guidance report on implementation for schools (Sharples, Albers and Fraser, 2018). This indicates that the context within which initiatives take place is not static but dynamic, and, furthermore, sometimes agentic: if we treat the support of senior school leaders as a contextual feature, senior leaders themselves will be undertaking other activities alongside the initiative being evaluated towards the same ends, of course.

Secondly, this indicates that contextual factors such as senior leader and organisational support are relational. They can act in different ways in relation to the programme at hand: as moderating influences; as independent agents of change; and as potential outcomes of the intervention. Furthermore, they can act at different points in the implementation process and can act in concert with or against other contextual factors.

Thirdly, the dynamic change processes that are undertaken within the settings can be over very long time periods. Coldwell (2016) notes, for example, literature on teacher development identifies that teachers develop their identities over long periods of time, moving through what Day and Gu (2010) call 'professional life phases', and their attitudes and responses to programmes are likely to vary in relation to these.

Finally, context is not external, it is internal to the decision-making of those involved. As Astbury and Leeuw (2010, p.370) argue, interventions "*work through human agents who have the (cognitive) capacity to think and act in terms of causalities and who also possess other capacities that make things happen. In practical terms, people do not react to programs like billiard balls that are hit* ". Coldwell (in press) links this to Bourdieu's concept of habitus: "*a 'practical sense' that inclines agents to act and react in specific situations in a manner that is not always calculated and that is not simply a question of conscious obedience to rules. Rather, it is a set of dispositions which generate practices and perceptions. The habitus is the result of a long process of inculcation, beginning in early childhood, which becomes a 'second sense' or second nature.*" (Bourdieu, 1993, p.5), suggesting that," f*or teachers, this might mean that their educational world has been so orientated towards using a particular pedagogical approach that the possibility of changing it in response to a professional development experience such as a training course would require such a shift in world view as to be almost impossible for them (which Ball (2003) argues can occur for teachers that have spent their entire careers working within a 'performativity'-driven system). For others, their habitus might mean the time is right for such a change to occur*." Thus the context for the intervention is integrated with the responses to the intervention in the decisions taken by those individuals engaged in the intervention: so the context is immanent rather than external.

To summarise, context can be dynamic, changing shape over time; agentic, creating not simply moderating change; relational, acting both as context for and outcome of the work of initiatives; historically located, involving change processes over a much longer period than the intervention being evaluated; and immanent, acting through - and as an intrinsic part of - participants' responses to the programme, not external to it.

Turning to complexity, Rogers (2008, p.36) identifies that more complicated logic models can deal with two more complicated designs. Firstly, logic models can be modified in visual form (using multiple pathways) to cope with interventions that include multiple causal mechanisms - the example given is a maternal and child health service program to both develop confidence in

parenting and encourage parents to adopt healthier nutrition. Secondly, via a similar approach, they can be modified where there are alternative causal mechanisms that may be enacted in different contexts - in other words the same intervention might work in different ways in different contexts to lead to the same ends.

However many of the features of complexity identified by Walton (2016) including non-linearity; emergence; adaptation; uncertainty and coevolution are difficult for logic models to manage. Rogers (2008) discuses complexity in relation to recursive causality - where feedback loops occur between the occurrence of higher and lower order outcomes - and suggests these can be dealt with by circular rather uni-directional models. Tipping points, Rogers suggests, can only really be dealt with via annotation of the model. Emergence requires a theory of change approach (Connell and Kubisch, 1998) that may include aspects of logic models, but these need to be iterative and therefore are not of use in relation to the kinds of interventions we will go on to discuss here. Overall, then, logic models can struggle to deal with complexity.

Summary: four issues with using logic models for evaluation designs

The discussion above indicates four issues that need to be addressed for logic models to be improved in relation to their usefulness in evaluation designs.

Firstly, they can helpfully lay out the implementation logic behind implementation pathways from development to outcomes; however attention must be paid to alternative and interacting implementation pathways.

Secondly, they require attention to be paid to causal mechanisms. Often causal mechanisms are ignored, and this can mean that it is not possible to distinguish between 'implementation failure' and 'theory failure' and the critiques above indicate that alternative and complementary causal mechanisms should be considered. In addition, in educational interventions there are often multiple independent inter-related causal processes involved in the same programme.

Thirdly, the role of the context within which evaluations play out is crucial, and under-researched. Context can be missing entirely from such models. As Pawson and Tilley (1997) remind us, interventions always and only take place in context. This is the force behind their exhortation to consider 'context mechanism outcome' combinations. In this sense logic models can miss the fact that mechanisms do not always 'fire' in Pawson and Tilley's language. Borrowing from critical realism, and alternative way of expressing this is to say they do not recognise that the causal powers of some such mechanisms are not always enacted - yet they still have these powers. Further, logic models can oversimplify context in a number of ways, by failing to capture that context can be dynamic, agentic, relational, historically located and immanent.

Finally simple path models indicate a two dimensional ontological understanding of causation; other perspectives [critical realism; complexity/systems theory] that are interested in causality assume layered, complex generative mechanisms that create outcomes observable in the social world that derive from interactions between structures and activities at different system levels and levels of social reality. Whilst logic models focus on the specific intervention and try to lay out clear causal processes involved, deeper, underlying social processes will be ignored, and important sets of

circumstances and relationships which are simply not amenable to a logic model approach can be simply treated as context.

In the next section, we discuss how we have aimed to use logic model-based approaches that try to deal with these issues - in differing ways and to differing degrees of success - in two studies in education to underpin evaluation designs.

Responding to the issues: Developing and using evidence-informed logic models

Whilst it is implied in the criticisms of experimental designs, the use of combinations of methods informed by logic models is sometimes noted but largely unexplored in the theory-based evaluation field. Rogers and Weiss (2007, p.65), for example, suggest combining theory-based causal models, that use a variety of methods with impact evaluations using experimental methods to uncover variation in relation two key issues noted in the previous section - "different levels of implementation and different contexts" , and this aligns with guidance on the role of Implementation and Process Evaluation (IPE) in EEF trials (Humphrey et al, 2015). However, the ways in which methods can be combined is not addressed by Rogers and Weiss (2007).

Just as the literature on theory-based evaluation is limited in its reference to combining methods, the methodological literature on mixed methods rarely discusses logic models. One notable exception is Yin's work in relation to Case Study. Yin (2013, p.324) notes that mixed methods case study evaluations "frequently use logic models, initially to express the theoretical causal relationships between an intervention and its outcomes, and then to guide data collection on these same topics." as in other evaluations. He argues, from a perspective of the need for mixed methods case study designs to improve validity of inference, that "most evaluations collect data about the boxes [in visual logic models], but nearly no data about the arrows. Yet they represent the flow of transitional or causal conditions, showing or explaining how one event (box) might actually lead to another event (a second box) […] For logic models not having any transitional data, only a correlational analysis can be conducted, reducing the causal value (and validity) of the entire exercise. Future studies could again investigate ways of improving the use of logic models." echoing similar arguments about a lack of focus on causation as those discussed used above by Weiss (1997) and Astbury and Leeuw (2010).

The rest of this section contributes to filling this gap in the literature by describing how we developed and used an evidence-informed logic model approach to bridge methods in evaluation. The examples given illustrate our developing thinking about creating evidence-informed logic models, which lead in our final discussion to a more developed frame.

Building evidence-informed logic models

In early EEF evaluations, in common with other evaluators, we worked with logic model designs based on the simple Wisconsin model[1] which did not address the key issues laid out in the previous

[1] It is important to note that the Wisconsin model developers recognise that more complex models with multiple pathways and feedback loops can be used, presenting some visual examples (Taylor-Powell, Jones and Henert, 2003 p.109). However, the downloadable templates from the Wisconsin website, used in projects including the book-gifting example, are of the simple variety.

sections. For example, Figure 2 below shows a simplified logic model prepared by programme implementers for a book-gifting programme with some associated activities over the summer holidays that aimed to improve engagement with reading for pleasure across the transition from primary to secondary school (when there is a dip in their achievement over this period, according to earlier research evidence).

Figure 2 here

This model exemplifies a common feature of such simple models, which is that the implementation logic is simply articulated, as would be expected given that it is based on the programme implementers' understanding of what will happen as a result of the intervention. However, the causal theory is missing, and the context is treated as external and largely static; and alternative implementation paths are not provided. This meant that the mixed methods design was over-simplified and the learning from the study was limited: although we were able to provide some explanation of the reasons behind the impact findings (in this case, that they did not show impact) these focussed on the extent to which the programme was faithfully adopted, rather than the contextual differences between setting or indeed the applicability of the underlying causal theory.

To help deal with the shortcomings of this kind of model, we began to develop what we call an **evidence-informed logic model** approach. We refer to our approach as evidence-informed for two reasons. Firstly, the model itself draws on research evidence presented above on the limitations of typical logic models in relation to theorisation of implementation logic, causal mechanisms, context and complexity. Secondly, as we will demonstrate in the next sections, the model is predicated on careful consideration of prior research evidence into the specific initiative being evaluated including its likely causal and implementation processes, in context and where appropriate taking into account complexity. As part of this second element we specifically look at the evidence relating both to the logical connections between elements in implementation paths and to the causal mechanisms that theorise how these connections occur and lead to sought for outcomes.

The two studies discussed in the next sections have been chosen as exemplars of mixed methods designs based on evidence-informed logic models. It is important to note that they represent the development of our thinking, and later in the paper we reflect on the extent to which they addressed the issues outlined earlier, enabling us to move forward to a final proposal for an evidence-informed logic model frame. The two studies have been selected to cover different methodological approaches - a pilot evaluation and a scale-up campaign evaluation. They cover different substantive aspects of education - a professional development programme for early career teachers and a campaign to improve deployment of Teaching Assistants (TAs). It should be noted, however, that both focus on primary rather than secondary education, and both are evaluations funded by the EEF. We selected EEF studies since, as noted above, it encourages the use of logic models and is the largest non-governmental sponsor of educational evaluation in the UK currently. Although neither example used in the paper involved a trial, such an approach is appropriate for RCT designs: and in fact we are using an evidence-informed logic model to underpin the mixed methods design of a current EEF trial of an approach to support functional linguistics in primary schools, 'Integrating English' (Sheffield Institute of Education, 2017).

Since the Integrating English trial was underway but not yet complete at the time of writing, we will provide just a brief outline here before turning to the other studies in more detail. Integrating

English consists of a programme of training for primary school teachers of pupils aged 9-11 to help them link teaching of grammar explicitly to the social and curriculum context of learners via the development of a scheme of work which is then implemented in classroom practice. In a similar way to the RETAIN model discussed below, the evidence-informed logic model considered two separate but linked causal processes - the process by which the training was expected to lead to practice changes, and the process by which the changed teacher practices were expected to lead to changes in pupils' outcomes. The logic model underpinned the design, firstly by helping to lay out the primary and secondary outcomes of the trial (in this case, impact of the intervention on the language ability of pupils in Year 6 aged 10 and 11 measured by a writing test being the primary outcome and by two other tests, of reading and grammar, punctuation and spelling, as secondary outcomes). Secondly, the model helped underpin the choice of fidelity measures to be incorporated into an 'on treatment' analysis. Thirdly, it helped with the development of a set of methods as part of the IPE design including surveys of teachers, observation of training and observation of classroom practice. This combination of approaches bridged by an evidence-informed logic model provides a strong basis for the final analysis, in the same way as RETAIN and the TA Scale-Up Campaign presented below.

Turning to the two studies, the South and West Yorkshire 'TA Scale-Up Campaign' was the first EEF 'scale-up of research-use campaign', designed to encourage schools to adopt practices that align with EEF guidance on the best use of Teaching Assistants (Sharples, Webster and Blatchford, 2015[2]) via a set of methods, which we discuss below. The evaluation (Maxwell et al, 2018a) utilised pre- and post-campaign surveys of all primary schools in South and West Yorkshire and a post-campaign survey of comparison schools (response rates 30-36%), case studies of participating schools, interviews with delivery partners and analysis of attendance data, with a separate impact evaluation based on a synthetic control that was to be undertaken by the Institute of Fiscal Studies in autumn 2018..

The 'RETAIN' early career teacher (ECT) Continuing Professional Development (CPD) programme evaluation (Maxwell et al, 2018b) was a pilot evaluation based in South West England which had a longitudinal design utilising repeated surveys of ECTs, semi-structured interviews with participating teachers, their in-school champions and head teacher and the delivery team, and observations of the programme.

In the section that follows, for each study, we demonstrate how we created a logic model to describe the intervention and underpin the mixed methods evaluation design, that takes into account the complex, situated nature of change processes and the perceived mechanisms and sequences that lead to that change. In each case, we lay out both the final visual representation and the methods used, showing how such a model helps in practice with building appropriate, sequenced evaluation designs using appropriate combinations of methods to meet the research aims, highlighting how this relates to the four core issues identified above (implementation logic, causal mechanism, context and complexity). In the subsequent section, we illustrate the ways in which constructing an evaluation design in this way increases the plausibility of claims that can be

[2] The guidance report contained seven recommendations for school leaders on the deployment of Teaching Assistants, for example, using TAs to supplement what teachers do, not replace them; ensuring TAs are fully prepared for their role in the classroom; using TAs to deliver high-quality, one-to-one and small-group support using structured interventions.

made from the study, again relating these to the four core issues. We also note the limitations of the logic model formats used in these evaluations.

Working with evidence-informed logic models in mixed methods evaluations: building the models

In both studies, the logic model was used to act as an explanatory framework which aimed to provide a theoretical model for the intervention. Below, we describe how each model was constructed taking into account - to varying degrees - the causal process including the implementation logic and causal mechanism/s, the role of context and the complexity of the social world, with the aim of better understanding the working of the intervention and interpretation of the impact findings.

In both logic models below (Figure 3 and Figure 4), inputs are the material and human resources and the delivery methods that comprise the intervention, and outputs are the immediate results of the intervention such as the number of teachers and schools involved in a programme, which could be summarised as the intervention's reach and engagement. We differ from the Wisconsin model, here, which treats resources as inputs and the programme or intervention itself as outputs. By outcomes we mean the intended and unintended effects of an intervention. These can be in relation to individuals (including teachers, leaders, parents and pupils), teams and organisations (especially schools and groups within them, but also policymakers, funders), and the social and political environment (local and national policymakers in particular). Intermediate outcomes are those that act to mediate as part of the process between the intervention inputs and longer term outcomes and impacts. By impacts we mean the outcomes that are the explicit goal of the intervention, often related to pupil attainment in educational evaluation. Other longer term outcomes that are not classed as impact are outcomes, such as enhanced school culture and capacity, which in the longer term act to sustain intended impacts. Contextual factors and complexity have been described above. At this stage of development of the model, we used the term enabling characteristics to describe the features of an intervention that can explain how the intervention works to lead to outcomes and impacts including some input characteristics. This aligns with what we would now describe (see Figure 6 below) as the underlying causal mechanisms.

1. Building the TA Scale-Up Campaign logic model

The TA Scale-Up Campaign evaluation used a logic model to frame the evaluation design and the methods. It was constructed by reviewing the relevant research literature - in this case, relating to knowledge mobilisation and research use, in particular the role of knowledge intermediaries - alongside conversation with EEF and other stakeholders about the specifics of the intervention (Figure 3). This combination of prior research and knowledge of the specific initiative was particularly important in relation to the design of this study and the logic model. Figure 3 is a slightly adapted version of the logic model used to underpin the evaluation. It is identical to that agreed at the start of the campaign with stakeholders, except for a restatement of inputs to make clearer the two separate causal processes (see below).

Figure 3 here

a) Identifying causal processes

The first causal process combined two elements, the provision of the EEF TA guidance to every school in England and a national campaign to encourage its use. The evidence base on the use of research summaries, such as the EEF guidance, indicates this combination is important. There is little evidence that the provision of such summaries and guidance can increase research use on their own. However, they may do so in combination with support, such as the EEF national campaign including press releases and running local awareness-raising events. For example Langer, Tripney and Gough's (2016) scoping review of social science literature 'suggests that advocacy and awareness-raising campaigns are effective in supporting behavioural change' (p.2), by providing opportunity and motivation to engage with research evidence.

Casting this evidence in relation to the four key issues in this paper, the implementation logic involved a series of steps from provision of guidance on use of TAs; followed by a set of campaign promotional activities (such as press releases, articles in educational media) and events; which leads to changes in school practices. The suggested causal mechanism by which this occurs is that the combination of increased opportunity to access evidence (the guidance) with increased motivation to engage with it (the national campaign) which is theorised to lead to behaviour change (uptake of recommendations contained in the guidance).

The second process was quite distinct: this was the use of intermediaries - referred in the TA Scale-Up Campaign as 'Advocacy Providers' - in facilitating research use.. Emerging evidence relating to intermediaries (Cooper, 2010 & 2014; Lavis et al, 2006; Lomas, 2007; Sin, 2008) suggested elements of intermediary work that were likely to lead to change. These included a set of attributes k that were likely to support engagement with and use of research including effective communication and interpersonal skills; understanding of research methodology and ability to find and assess relevant research and communicate with researchers . Alongside this, the knowledge mobilisation literature (including Langer, Tripney and Gough, 2016; Nelson and O'Beirne, 2014; and Nutley, Walter and Davies, 2007) provided more detail on how this might work by a set of processes which we lay out in detail in the project report (Maxwell et al, 2018a). They can be summarised as focusing on supporting contextualised use of research evidence via building skills amongst school leadership; supporting communication; and enabling the transformation of research by combining it with practice-based knowledge in school contexts. This provided the bedrock for developing a clearer causal process for this aspect of the initiative.

Discussion with the EEF and Advocacy Providers enabled us to draw out an implementation logic. Advocacy providers engaged school leaders in a set of workshops with schools and provided ongoing support to make changes to TA deployment aligned with EEF guidance recommendations, via a school-based change process. The research evidence outlined above provides the causal mechanism/s by which this was expected to occur, which could be summarised as follows. Intermediaries provide contextualised access to evidence (translation of guidance to fit local context); they helped build motivation to utilise evidence (peer to peer support, effective communication by the providers) and provide improved capability to implement change (facilitating development of clear change processes in schools, supporting change management skills of in-school leaders).

The logic model hypothesises that, taken together, these processes would then enable the achievement of intermediate outcomes for pupils, such as enhanced knowledge, skills, engagement

and confidence. The final outcomes were hypothesised to be school practices aligned with the EEF guidance and increased engagement with research evidence, with a hope that in the longer term this would lead to improved pupil attainment, due to improved pedagogical and organisational practices.

b) Considering context and complexity

In relation to context, recent research evidence highlights the importance of the organisational context, particularly leadership capacity and commitment, and the impact of the educational policy context, in determining research use (Coldwell et al, 2017), indicating aspects of the context that needed to be considered. Additional potential contextual factors including those related to the advocacy providers such as their motivations and experiences of research use and prior relationships with schools were added following discussion with stakeholders.

Finally in relation to complexity, whilst the earliest models of knowledge mobilisation suggested a one-way 'transfer' of evidence from researchers to practitioners (more recent conceptualisations suggest complex and less linear knowledge flows (Powell et al, 2017), thus foregrounding both complexity and the need to be open to multiple implementation routes.

As can be seen from the presentation in Figure 3, the careful consideration of research evidence combined with discussion of the specific context is extremely difficult to represent in a typical visual logic model. Therefore our approach was to use the more detailed, in-depth understanding of the likely complex ways in which the programme might work as a resource sitting behind the simpler visual presentation. This kind of approach is advocated by Rogers (2008). Despite its limitations, the model did lead to insights beyond what might have been possible with a simpler underlying model as we will go on to discuss in the next section.

2. Building the RETAIN logic model

Similarly to the TA Scale-Up Campaign, the mixed methods design for the evaluation of RETAIN drew on a logic model (Figure 4) that was informed by prior evidence and discussion with EEF and - in particular - the delivery partners who also drew extensively on research evidence to underpin the design of the programme. As with the TA Scale-up campaign logic model, we were still using the term enabling characteristics to describe the features of the intervention that could potentially explain how the intervention works to lead to outcomes and impacts - which we now describe (see Figure 6 below) as causal mechanisms.

Fig 4 here

a) Identifying causal processes

Examination of the evidence enabled us to identify two related causal processes, the first of which has two branches. The first causal process suggests an implementation logic of a fairly complex CPD programme leading via improved teacher knowledge, skills and self-efficacy to changes in teacher practice as evidenced more widely in evaluation literature (see for example, Desimone, 2009). The causal mechanisms underpinning this implementation logic were particularly drawn from a range of reviews on effective features of CPD programmes (including Cordingley et al, 2015, Timperley et al, 2007, Yoon et al, 2007), which provide evidence for the effectiveness of some components of the

RETAIN programme in facilitating teacher change. These include peer collaboration, teacher ownership of their learning and engaging teachers in experimenting with their practices. Other evidence supported the impact of coaching, another core component of RETAIN (for example Kraft et al, 2016, found a large positive effect on quality of instruction), with research in the related field of mentoring providing some evidence to support the importance of external (as opposed to in-school) support (for example Cameron and Grant, 2017 and Hobson & McIntyre, 2013). One branch of this first causal processes concludes with teacher practice change, which in turn is the input to the second causal process, discussed below. The second branch of the first causal process branches off from changes in teacher self-efficacy and leads to teacher retention. There was some evidence, especially in relation to early career teaching, (for example: Ashby et al 2008, Day and Gu 2010, Coldwell 2017, Buchanan et al 2013) to support this implementation path although the evidence base lacked robust impact studies. Although the evidence is limited, there are some indications of a causal mechanisms that comes into play as higher teacher efficacy leads through increased satisfaction and positive orientation to increased intention to stay in teaching (Brouwers and Tomic (2000); Tschannen-Moran, Hoy & Hoy, 1998; Klassen, Tze, Betts & Gordon, 2011).

y. The second causal process, which follows sequentially from changes in teacher practices at the conclusion of the first branch of the first casual process, has an implementation logic whereby improved teacher practices , with underpinning pedagogical strategies that are informed by research evidence lead ultimately to pupil outcomes. The mechanisms underpinning this implementation logic include teacher's sense of self-efficacy impacting on a range of positive outcomes for pupils including cognitive achievements and attainment (see for example: Caprara et al; 2006; Muijs & Reynolds, 2001), mechanisms associated with coaching (Kraft, 2016) and potentially research evidence use (see discussion related to the TA project above).

.

b) Considering context and complexity

The research review did not reveal specific contextual features or complexity, but these emerged from conversations with project leads as indicated in the project logic model presented in Figure 4.

The multiple-inter-related components of the RETAIN programme, taken together with the multiple and differing aims of the programme introduced a high level of complexity. The inter-related nature of the two branches of the first causal process and the overlaps and relationship between the second and first causal process is clearly evident from the discussion above. This presented significant challenges for developing the logic model-based. Our approach was to unpick the differing elements of the programme and their hypothesised causal processes, and examine the evidence behind each before linking these together to assemble an overarching model.

As with the TA Scale Up Campaign design, the limitations of a visual model in capturing this complexity is clear, but this model has some improvements in that it allows back and forth linkages to be represented; it deals with differing types of outcomes; and - importantly it shows multiple, interrelated independent causal processes in relation to effective professional learning on the one

hand and processes related to how this learning can lead to practice, retention and pupil outcome changes on the other.

In the next section, we discuss how these logic models were used to build evaluation designs combining a set of appropriate methods, and the benefits that accrued from them.

Working with evidence-informed logic models to build evaluation designs combining methods

Using the TA Scale-Up Campaign logic model to bridge methods in the evaluation

The TA Scale-Up Campaign design drew on the logic model (Figure 3) to develop an implementation and process evaluation design including a set of inter-related elements as follows.

In relation to the first ('campaign') causal process, repeat interviews were undertaken with the EEF implementation team, to help understand the implementation of the campaign at national level.

In relation to the second ('advocacy') causal process, methods included:

- Repeat interviews with advocacy providers.
- Observations of EEF-led advocacy provider meetings.
- Case studies of participating schools after campaign completion.

These methods, especially the case studies, also provided evidence relating to contextual variation, and complexity.

These were linked with the following methods, examining intermediate outcomes from both processes as they worked together, in line with the logic model:

- Pre- and post-campaign surveys of participating and non-participating schools in South and West Yorkshire.
- Post-campaign survey of a group of comparison schools.
- Analysis of data on the attendance of school staff at workshops and training events, and of participation in structured TA-led intervention randomised controlled trials (RCTs).

Longer term impact in relation to pupil attainment using a synthetic control were due to applied in Autumn 2018.

This set of methods demonstrates how the logic model allowed us to direct resource into data gathering in ways that may not have been apparent without its use. Firstly, it allowed the evaluation team to build samples that represented the key stakeholders in the programme that may not all have been targeted without developing the model. In particular, the role of EEF not simply as a funder but as an actor in the programme in a variety of ways (for example, as provider of the guidance report and as a challenge and support to schoolsand Advocacy Providers) emerged at the logic model development stage and suggested the need for focussing data gathering on EEF as well as Advocacy Providers and schools/participants.

Secondly, it focused this data gathering, including the detail design of research instruments to actively look for whether the implementation path was proceeding as expected, and it allowed us to focus on the assumed causal mechanisms. Thirdly, it allowed us to focus on the role of the

contextual features both of the differing forms of Advocacy Provider and the settings (local authority and school) within which they tool place. Finally, it allowed connections to be built between elements of the model.

This enabled exploration of how the campaign was implemented in South and West Yorkshire, to what extent schools were supported by the advocacy provision, participants' experiences of, and responses to, this support and how, and in what circumstances, this led to positive changes in the use and practices of TAs in schools. The agreed focus for impact in the process and implementation evaluation was on the intermediate outcome of school practices. Measurements of impact on practice in the survey and questions to explore practice change in school case studies were developed using the seven EEF recommendations for the effective deployment of TAs.

In addition to these benefits - associated with selection of respondents, method and design and content of data collection methods - the logic model allowed us to develop a stronger analysis of the impact of the campaign and Advocacy Provider approach and an associated analysis of how such approaches can be developed in the future. Drawing on the mixed methods developed on the basis of the logic model, data collected and analysed provided evidence of impact on practices (for example, there was some evidence that practices in participating schools were more closely aligned to the EEF recommendations at the end of the campaign than the beginning**,** although there was very little evidence to indicate any difference between the practices of participating schools in South and West Yorkshire and the comparison group schools; however the lack of pre-test measures in comparator schools meant this judgment was not secure). Qualitative data collection indicated that the campaign appears to have led to sustainable change in schools. Overall, the data collected suggested that Advocacy Provision could be a useful strategy that can support EEF's remit to increase research use.

Overall, we were able to judge that evidence from the evaluation broadly supports the assumed implementation logics set out in the initial project logic model and provided deeper insights into the change processes including the causal mechanisms behind and contextual variations associated with the logic model - for example, that implementation of the recommendations at the school level broadly followed a sequence of recognition of problem and need for change, data gathering, project design and planning, implementation and review.

The logic model and its focus on the differing roles of those involved allowed the analysis to articulate the causal processes including implementation logic and causal mechanisms. So, for example, the Advocacy Providers worked to convene and provide structure for change processes in schools; schools themselves worked to create the change most effectively where there was leadership commitment, a culture of commitment to improve outcomes for all and strong relationships between TAs and teachers. Analysis revealed that the role of EEF as provider of the guidance but also significantly as a broker and national advocate was important.

The focus on the different elements of the expected causal processes alongside the treatment of context as dynamic and active, allowed analysis to lead to Figure 5, a simplified presentation of the relationship between the two causal processes and one crucial contextual feature designed to support future implementation. In this model the 'campaign' causal process is represented by the research object being used – in this case the EEF guidance and recommendations, and national work to encourage take up. The 'advocacy' causal process relates to the advocacy providers and provision.

The third elements relates to schools as the crucial contextual component. The inter-relationship between the three - a different presentation than in a standard logic model - represents an element of the complexity involved.

Figure 5 here

A key learning point from the evaluation was that that there were sets of characteristics that needed to be in place in all three components of this diagram for effective advocacy to occur. So, for example, characteristics of the first causal process relating to the research object included that it was provided by a trusted provider (EEF), and was user-friendly and based on robust research. Characteristics of the causal process relating to the advocacy included advocacy providers being professionally credible, knowledgeable individuals able to provide challenge as well as support, using a process that provided a clear framework that enabled schools to sequence change and implement manageable steps. Characteristics of schools as crucial contextual features included motivated staff and committed leaders.

This analysis, then, indicated that, in similar projects, attention needs to be paid not just to the research object/campaign and the provision itself, but to the engagement and characteristics of schools. By considering these as separate, inter-related system components each of which has causal power, the logic model approach used allowed for this insight to emerge at the analysis stage.

Using the RETAIN logic model to bridge methods in the evaluation

Turning now to RETAIN, the approach used showed the particular value of a logic model approach for pilot evaluation. Given that, as a small-scale, short term evaluation, the intended programme outcomes of raising attainment for disadvantaged pupils and increasing teacher retention could not be directly measured, "evidence of promise" was assessed using a set of measures. The measures related to the outputs and intermediate outcomes set out in the logic model with the underpinning assumption that if these outputs and intermediate outcomes are achieved, these will lead over time to the intended final outcomes. The evaluation was strengthened by selecting methods and designing instruments that tested the plausibility of the proposed casual processes, in addition to developing and using key indicators linked to intermediate outcomes. This included designing a framework to assess alignment with research-informed principles underpinning effective CPD.

In contrast to the TA campaign the complex inter-relationships between the causal processes led to a common set of data collection methods being adopted to explore both these causal processes and intermediate outcomes. These methods were:

- Repeated surveys of early career teacher participants over the duration of the programme
- Participant interviews and focus groups over the duration of the programme
- Repeated interviews with the programme delivers and school informants - head teachers and the member of staff designated to act as a school champion for the RETAIN in participants' schools.

In addition, in relation to the first causal process, whereby an effective CPD programme leads via improved teacher knowledge, skills and self-efficacy to changes in practice, observation of regional workshops and videos of taught sessions were undertaken alongside a professional review of

programme documentation and resources. Data were also collected from interviews to explore contextual variation and the dynamic impact of context on causal processes.

The logic model was important in ensuring that the design of each research instrument and subsequent analysis of data fully explored each causal process. Firstly, in relation to the first branch of the first causal process associated with CPD influencing teacher's practice, a framework of indicators of effective CPD was constructed from research evidence. This framework adapted and extended Desimone's (2009) set of critical features of effective CPD to include: content, active learning, relevance, collaborative learning, duration and rhythm and sense of shared purpose about professional development. Drawing particularly on Cordingley et al's (2015) review of international reviews of effective professional development we populated the framework with indicative criteria for each characteristic. The framework and indicative criteria where shared with ECTs, school champions and the RETAIN team and their perceptions of the extent to which the criteria were met were collected though the interviews. Head teachers' perceptions of the extent to which the RETAIN programme had met the criteria related to a shared sense of purpose for professional development between the school and individual teachers were also collected through interview. The framework was then populated with key findings from a thematic analysis of the interview data, analysis of the evaluators' observation of regional workshops and videos of taught session field notes and the professional review of programme documentation and resources.

Secondly, also relating to the first causal process, whereby CPD can lead to practice changes, likert scale self-report survey questions relating to ECTs' knowledge and skills, confidence, awareness and use of research evidence and practice were developed. Alongside this, the perceived relationships between the components of RETAIN individually and in combination were explored with all interviewees.

Data were gathered via interviews to illuminate the second branch of the first causal process whereby improved self-efficacy and other outcomes of the CPD programme were intended to lead to teacher retention. Inevitably, given the relatively short-time frame for the pilot, these data were limited and impressionistic. In relation to the second causal process, whereby improved teacher practices can lead ultimately to pupil outcomes via mechanisms associated with coaching and potentially research evidence, the evaluation was again reliant on the perceptions of interviewees.

The centrality of teacher self-efficacy in both causal processes was highlighted though the development of the evidence-based logic model, as indicated above. It therefore became evident that incorporating a validated measure of teacher self-efficacy (the Teachers' Sense of Efficacy Scale (TSES), Tschannen-Moran & Hoy, 2001) into the evaluation would provide evidence in relation to a key intermediate outcome which, if positive, would provide some support to claims of evidence of promise in relation to the intended final outcomes of both causal processes. Such an approach has the potential to help to address some of the limitations in in evaluations that have a relatively short time-span, bearing in mind the small-scale nature of this study and the lack of any comparator meant that that the design was limited. Nonetheless, the RETAIN evaluation has demonstrated the utility of using the TSES in evaluations that include a causal process related to professional development.

Similarly, the framework for assessing effective CPDs developed for this evaluation outlined above, also has potential utility in evaluations that include professional learning. These tools were both

derived from the evidence-informed logic model and could be further tested in longer term studies to better understand their predictive utility in relation to final outcomes occurring.

Use of the evidence-based logic model to inform the evaluation design enabled a sufficient range of data to be collected, as well as providing frameworks for analysis. This enabled defendable judgements to be made in relation to changes in ECT practices, professional learning, and career development and other intended intermediate outcomes. Teachers and leaders were, in some cases, able to articulate examples of how changes had occurred in line with the change process predicted by the logic model and in particular how the combined influences of three of the six components of RETAIN, namely participating in 1) the taught sessions underpinned by research evidence, 2) coaching and 3) peer collaboration had led to positive intermediate outcomes. For pilot studies such as the RETAIN evaluation that are not able to use comparators, articulating and providing evidence of the path from inputs to outcomes described in the causal processes in an evidence-based logic model helps support the plausibility of the model, and can support the potential of an approach for testing via an efficacy trial.

The evidence-informed logic model approach indicated suggested improvements to both strengthen the hypothesised links between programme inputs and intended outcomes and increase the attractiveness of the programme to ECTs and schools, and illuminated a key issue for RETAIN, and other CPD programmes, in that the effectiveness of the CPD approach in terms of practice change and ultimately pupil outcomes appears to be limited when schools are not open to change and/or are very prescriptive about teaching and learning approaches and resources.

Discussion

The evidence-informed logic model approach allows sequenced causal change processes to be fully theorised in the design phase of evaluations if they take into account the complexity and situated nature of educational change processes. The examples of the development of designs drawing on evidence-informed logic models presented above show both how such approaches can be undertaken, and the potential benefits.

The TA Scale-up Campaign evaluation showed how we were able to use consideration of prior evidence on the different aspects of the programme (on knowledge mobilisation in general; on national campaigns; on research summaries; on the role of research intermediaries) to develop a logic model with two related causal processes, related to the campaign and the advocacy provision, drawing out the implementation logic and potential causal mechanisms for each. This theorised how these elements were expected to work to lead to sought for outcomes via a set of potential implementation pathways. Further, we explicitly sought out evidence on the contextual factors likely to influence the success of the programme, and any underlying issues related to complexity (in this case, non-linearity). This approach bore fruit in a number of ways. In the previous sections, we highlight that it enabled the evaluation team to select appropriate respondents, using a range of methods that focussed on the expected causal process, contextual issues and potential complexities to gather data. At the analysis stage, this allowed consideration of the role of different system features - the research object, the school and the role of research intermediaries - acting together to lead to change.

Similarly, the RETAIN evaluation design highlighted how, in complex intervention designs, prior research evidence on different aspects of change processes is important. In this case, evidence related to effective strategies for engaging teachers in research-informed thinking about teaching for improved learning and increased pupil attainment, effective professional learning and the relationship between professional learning and retention of teachers in the profession. This enabled the evaluation team to separate out two inter-related but distinct causal processes. The first was a CPD programme leading via improved teacher knowledge, skills and self-efficacy to changes in practice, with a branch from self-efficacy to improved satisfaction and positive orientation towards the profession and subsequently greater retention in the profession. The second causal process traced the links between research-informed teacher practices to improved pupil outcomes. For each causal process the implementation logic and causal mechanisms were drawn out. This approach enabled the development of an evaluation design that took account of the complexity of the RETAIN programme design (the individual components and how they acted in combination) and the inter-relationships between the different causal processes. The complex inter-relationship between the causal processes is illustrated in our choice of presentation of two causal processes with one branching at the intermediate outcome stage, as this could equally justifiably be presented as three distinct causal processes. Evidence of potential contextual factors, to be explored through the evaluation was also identified from the literature and built into the evaluation design. The evidence-based logic model approach also supported the choice of existing measures and the development of new ones to provide evidence of promise for the programme. The resulting research design allowed data collection using a range of methods and subsequent analysis to convincingly examine the plausibility of the RETAIN programme evidence-informed logic model. In turn this enabled the evaluators to identify three core components that acting together are essential to positive outcomes (evidence-informed taught sessions, in-school-coaching and peer collaboration). In addition, the team could suggest improvement, and recommend that RETAIN was suitable for efficacy trial. A further outcome was the development of a qualitative framework for examining effective CPD in future pilot evaluations with CPD components.

More generally, the evidence-informed logic models in each case enabled appropriate methods to be used to examine both the causal processes and their relationship to programme outcomes, by means such as utilising teacher surveys, case study and interview approaches. In addition, the approach allows these methods to test out the contextual factors that enable the change process to be enacted, taking into account the complexity of the social world by - for example - showing that changes to teachers' use of research findings requires a focus on the research object; the change process; and the school environment as indicated in Figure 3. The examples used in this paper demonstrate the importance of situating the model in its specific empirical context; explicitly laying out the causal mechanisms and implementation logics, linking as appropriate to wider social theory; and drawing on prior research literature in combination with the insights and perspectives of funders and deliverers. They thus allowed us to strengthen our learning from the evaluations by allowing the capture of data on decision-making via individuals; to look for variety; and to recognise complexity to help in not over-claiming.

Reflecting on the models used in these examples (Figure 3 and Figure 4) indicates limitations that should be addressed. Firstly, the differing causal processes - in relation both to the implementation logic of each, and the causal mechanisms - are not clearly delineated. The causal mechanisms, in particular, are presented as part of a set of what we called enabling characteristics which combined

features of inputs and causal mechanisms, which we now suggest need to be separated. Further, the expected relationships between different causal processes are not always clear in the Figure 3 and 4 visual representations. Finally, the potential ways in which aspects of context may influence different causal processes at different points in the implementation process, according to the implementation logic, and the possibility of some contextual factors acting both as outcome and context, are not clear. To help overcome these limitations, a new visual logic model frame is proposed in Figure 6.

Figure 6 here

Figure 6 includes two separate but linked causal processes, which are represented by two distinct implementation logics, and two separate descriptions of the expected causal mechanisms. The contextual factors that may be included in the context box may influence different points in each implementation path, according to the separate implementation logic of each, and there may be some such factors that operate as both context and outcomes as indicated by double headed arrows.

In Figure 6, the two implementation logics are linked at the point at which the first influences the second. This kind of example is useful for interventions that include a causal process associated with teacher change and a further causal process associated with teacher change leading to pupil outcomes, for example. However, it is possible that more than two causal processes are needed, and the relationships between them may differ (as with the RETAIN model). In these cases, the model would need to be adapted. It is highly likely that aspects of the model - in particular the detail of the evidence underlying the causal mechanisms and the implementation logic behind the causal processes - will require further annotation to sit behind this visual representation. It is important to note that for evaluation studies involving approaches that may have complex features, like adaptation and system inter-relationships, the implications of this complexity will necessarily only become clear, and therefore can only be dealt with, at analysis stage.

Having presented this new evidence-informed logic model frame, we now turn to a consideration of their potential use in practice. Evidence-informed logic models can bring about helpful discussions and uncover complexity, but like any other models they are partial and have limitations. The issues raised in the paragraph above relating to complexity, in particular, help to clarify a salient point about logic models, or any other frameworks for research design for that matter: they can be helpful, but should not be reified as if they are faithful representations of the world. In other work discussing a set of models related to professional learning (Boylan et al, 2017), we argued that such models are best thought of as tools as part of project and evaluation design rather than as an endpoint. Some of the models discussed in this earlier work, particularly Guskey's (2002) and Desimone's (2009) path models are very clearly related to logic models, and we would argue therefore that treating logic models as tools is also appropriate. In the final part of this discussion we suggest how logic models can be used as tools to support evaluation in practice.

Table 1 below presents one way of doing so, by presenting a set of questions that can help shape the development of an evidence-informed logic model drawing on the kind of frame presented in Figure 6. Note some of these are developed from Coldwell (in press).

Table 1 here

Once an initial model has been developed, drawing on Table 1 and Figure 6, an evaluation design using a set of appropriate methods can be developed and undertaken. The previous section highlights at least four ways in which the model can help frame and support such an evaluation.

Firstly, it can help direct data-gathering towards particular sets of respondents, since the model will help identify which groups, organisations and individuals will be involved. This could be in one or more of the following ways. They might be engaged as (potential or actual) direct recipients of the impact of the intervention (often pupils); as comparators or controls; as recipients of intermediate or other longer term outcomes (often teachers, leaders, schools); they might be involved in the delivery of the intervention (for example designers; deliverers; funders); and they might be important in relation to understanding the influence of the context (for example, as part of the wider school, political, social or economic context).

Secondly, it can help decisions to be made about the methods to be used. There are always choices available here, but the implementation logic and expected causal mechanisms involved may suggest particular methods such as observation of training as part of implementation, or observation of teaching to ascertain if teaching practice has changed in line with the expected causal process. Similarly, the expected contextual factors may suggest methods such as the use of multiple perspectives via school case study if the wider school context is seen to be significant. As noted above, emergence adaptation and feedback loops - and other features of complexity - may only become apparent as the intervention takes place so it is less likely that complexity will influence choice of methods.

The third way in which an evidence-informed logic model can help is in the content focus of the methods, as exemplified by the design and focus of survey and other tools, for example the framework for assessing the effectiveness of CPD as described in relation to RETAIN above.

Finally, and perhaps most clearly in terms of bridging these methods, the model can help in the analytical phase. Implementation and context factors can be included in statistical modelling, of course, and data analysed as part of implementation and process evaluations can support interpretation of variation in impact in experimental or quasi-experimental designs. Furthermore, as indicated above, the logic model performs a useful function in laying out expected causal processes which - as illustrated above in relation to RETAIN in particular - can help test whether the expected implementation paths underpinned by the implementation logic play out in practice, and whether there is evidence that the theorised causal mechanisms are enacted as part of this process. While this has particular pertinence in pilot evaluations when other methods of measuring impact are inappropriate, it can also play an important role in trials, for example in explaining null effects. For other evaluation designs, such as the TA Scale Up evaluation, it can help with subsequent theorisation of causal processes that may involve more complexity, as indicated by the presentation in Figure 5 and associated discussion.

Conclusion

Whilst the approach to building and using evidence-informed logic models described here can improve their use in educational evaluation, the difficulties such models have in dealing with

complexity indicates that they should be treated as one possible approach, which is appropriate in some circumstances but is not any kind of panacea for educational evaluations . They are most useful for clearly specified interventions, but of course other alternatives are available such as Realist 'context-mechanism-outcome' combinations (Pawson and Tilly, 1997). Further, they can be inappropriate especially for highly complex programmes, and here - as Connell and Kubish (1998) argue - 'theory of change' approaches may have more traction. And as Biesta (2010) reminds us, such models can say nothing about what *ought* to be done. For example, initiatives with a focus on school cultural change must recognise it is a complex multifaceted process, with multiple change processes over differing time scales and at different system levels working together and in opposition, alongside deep social and economic change. A misrecognition of this by focussing on a specific logic model can lead to false claims about school effectiveness. For example, this must be the case where it comes to English inspectorate judgments which appear to rely on a form of simple logic from leadership to teaching to pupil results with the result that some schools are systematically seen to be failing. Systematic failure is likely to have at least some causes well beyond schooling and to attribute this failing to poor leadership or teaching per se is going to contribute to faulty causal thinking.

This complexity also highlights a particular limitation of evaluation approaches that focus on testing interventions via randomised controlled trials. Such evaluations are far more straightforward for interventions that are bounded and codifiable, with clear definable outcomes to make it possible to randomise, to avoid spill over effects and so on. This can funnel research efforts into evaluation of initiatives that are more amenable to trial, and so can contribute to focussing research and policy efforts on interventions with simple causal processes, rather than focussing attention on more complex change processes that may or may not be located in the school which even an evidence-informed logic model approach such as that advocated in this paper may struggle to deal with.

Bearing in mind these limitations, this paper has aimed to show how evidence-informed logic models can helpfully bridge methods in some evaluation designs. In an insightful discussion of the state of the field in the late-2000s, Stame (2010) argues that evaluation approaches have been split between a focus on internal validity, aiming for the strongest possible evidence of causation, largely from within the experimental tradition, and a focus on external validity, aiming to provide the most useful transferable learning, largely from the theory-based tradition. This paper suggests that an evidence-informed logic model approach cannot not only bridge methods but also provide a bridge between these traditions, by using careful consideration of research evidence to build models that help lay out the plausible pathways leading to change based on causal processes that include identifying implementation logic and casual mechanisms, linked together with associated impact evaluations.

Earlier in the paper, we noted that one argument made about the use of overly simplistic logic models is that they confuse implementation logic and causal mechanisms, and that by separating out these two elements of the model, as in Figure 6, evaluators are better able to distinguish implementation failure from theory failure. Given the epistemological limitations of social research, we prefer the terms implementation deficiency and theory deficiency, and with Lipsey et al (1985) we would add a third: methodological deficiency (or failure as Lipsey would have it). Whilst Lipsey - working within the experimental tradition - was focussed on lack of statistical power, we would argue that using approaches drawing on evidence-informed logic models that pay attention to the

issues raised in this paper can mitigate methodological deficiency or failure in a broader sense. Designs that use evidence-informed logic models are able to more fully consider causation, implementation logics, context and complexity than simple logic model approaches that tend to be used at the moment. In this paper, we have presented evidence of how the approach used in two such designs can lead to stronger analysis and more secure judgments about the workings of programme at hand. Further, this can help provide stronger theorisation of empirical work to support future work in the field.

References

Ashby, P, Hobson, A J, Tracey, L, Malderez, A, Tomlinson, P D, Roper, T,... & Healy, J (2008) *Beginner teachers' experiences of initial teacher preparation, induction and early professional development: A review of literature*. London: DCSF.

Astbury, B and Leeuw, FL (2010) Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation* 31(3): 363-381.

Ball, S J (2003) The teacher's soul and the terrors of performativity. *Journal of Education Policy* 18*(*2): 215-228.

Barber, M, & Mourshed, M (2007) *How the world's best-performing schools systems come out on top.* McKinsey & Company.

Bickman, L (1996) The application of program theory to the evaluation of a managed mental health care system. *Evaluation and program planning* 19(2): 111-119.

Biesta, G J (2010) Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in philosophy and education* 29(5): 491-503.

Bourdieu, P(1993) *Sociology in question* (Vol. 18) London: Sage.

Boylan, M, Coldwell, M, Maxwell, B, & Jordan, J (2018).Rethinking models of professional learning as tools: a conceptual analysis to inform research and practice. *Professional development in education* 44(1): 120-139.

Brouwers, A, & Tomic, W (2000) A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher education.* 16(2): 239-253.

Buchanan, J, Prescott, A, Schuck, S, Aubusson, P & Burke, P (2013) Teacher retention and attrition: Views of early career teachers. *Australian Journal of Teacher Education* 38(3): 112-129.

Cameron, D & Grant, A (2017) The role of mentoring in early career physics teachers' professional identity construction", *International Journal of Mentoring and Coaching in Education* 6(2): 128-142.

Caprara, G V, Barbaranelli, C, Steca, P, & Malone, P S (2006) Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of school psychology* 44(6): 473-490.

Chen, H-T (1990) *Theory-driven evaluations.* Newbury Park, CA: Sage.

Coldwell, M (2016) Career orientations and career cultures : individual and organisational approaches to beginning teachers' careers. *Teachers and Teaching: Theory and Practice* 22 (5): 610-624.

Coldwell, M (2017) Exploring the influence of professional development on teacher careers: A path model approach. *Teaching and teacher education* 61: 189-198.

Coldwell, M (in press) Reconsidering context: six characteristics of context to improve learning for evaluation. *Evaluation*

Coldwell, M, Greany, T, Higgins, S, Brown, C, Maxwell, B, Stiell, B, Stoll, L, Willis, B & Burns, H (2017) *Evidence-informed teaching: an evaluation of progress in England.* London, DfE. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/625007/Evidence-informed_teaching_-_an_evaluation_of_progress_in_England.pdf

Connell, J and Kubisch, A (1998) Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems, in K Fulbright-Anderson, A Kubisch and J Connell (eds) *New Approaches to Evaluating Community Initiatives*, vol. 2, Theory, Measurement, and Analysis. Washington, DC: Aspen Institute.

Cooper, A (2010) *Knowledge mobilization intermediaries in education*. In: The Annual Conference of the Canadian Society for the Study of Education (CSSE): Montreal.

Cooper, A (2014) Knowledge mobilisation in education across Canada: A cross-case analysis of 44 research brokering organisations. *Evidence & Policy: A Journal of Research, Debate and Practice* 10(1): 29-59.

Cooper, A, & Shewchuk, S (2015) Knowledge Brokers in Education: How Intermediary Organizations Are Bridging the Gap between Research, Policy and Practice Internationally. *Education policy analysis archives* 23(118).

Cordingley, P, Higgins, S, Greany, T, Buckler, N, Coles-Jordan, D, Crisp, B,... & Coe, R (2015) *Developing great teaching: lessons from the international reviews into effective professional development.* London: Teacher Development Trust.

Cox, R J (2000) Using programme logic models in evaluation: A review of the literature and the spinal outreach team experience. *British Journal of Occupational Therapy* 63(3): 115-120.

Day, C and Gu, Q (2010) *The new lives of teachers*. London: Routledge.

Desimone, L (2009) Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher* 38(3): pp.181–199.

Desimone, L M, & Hill, K L (2017) Inside the Black Box: Examining Mediators and Moderators of a Middle School Science Intervention. *Educational Evaluation and Policy Analysis* 39(3): 511-536.

Greany, T, and 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:.147–170

Greene, J C (2005) *Context,* pp. 82–84. S Mathison (Ed.): Encyclopedia of evaluation. Thousand Oaks, CA: Sage.

Guskey, T R (2002) Professional development and teacher change. *Teachers and Teaching* 8(3): 381-391.

Hattie, J (2009) *Visible learning: A synthesis of over 800 meta-analyses related to achievement*, London: Routledge.

Hobson, A J, & McIntyre, J (2013) Teacher fabrication as an impediment to professional learning and development: the external mentor antidote. *Oxford Review of Education* 39(3): 345-365.

http://dx.doi.org/10.1787/9789264xxxxxx-en http://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

Humphrey, N, Lendrum, A, Ashworth, E, Frearson, K, Buck, R, & Kerr, K (2016) *Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature.* London: Education Endowment Foundation.

Klassen, R M, Tze, V M, Betts, S M, & Gordon, K A (2011) Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise?. *Educational Psychology Review* 23(1): 21-43.

Kraft, M A, Marinell, W H, & Shen-Wei Yee, D (2016) School organizational contexts, teacher turnover, and student achievement: Evidence from panel data. *American Educational Research Journal* 53(5): 1411-1449.

Künzle, B, Kolbe, M and Grote, G (2010) Ensuring patient safety through effective leadership behaviour: a literature review. *Safety Science* 48(1): pp.1-17.

Lacouture, A, Breton, E, Guichard, A, & Ridde, V (2015) The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implementation Science* 10(1): 153.

Langer, L, Tripney, J, & Gough, D A (2016) *The science of using science: researching the use of research evidence in decision-making*. London: UCL Institute of Education, EPPI-Centre.

Lipsey, M W (1993) Theory as method: Small theories of treatments. *New Directions for Evaluation,* 57: 5-38.

Lipsey, M W, Crosse, S, Dunkle, J, Pollard, J, & Stobart, G (1985) Evaluation: The state of the art and the sorry state of the science. *New Directions for Evaluation* 27: 7-28.

Maxwell, B, Willis, B, Culliney, M, Coldwell, M, Demack, S, Goepal, J & Stevens, A (2018a) *Improving the impact of Teaching Assistants : EEF scale-up campaign*. London: Education Endowment Foundation.

Maxwell, B, Clague, L, Byrne, E, Coldwell, M, Culliney, M & Hobson, A (2018b) Evaluation of the RETAIN Early Career KS1 CPD programme. London: Education Endowment Foundation.

McIntyre, J, & Hobson, A J (2016) Supporting beginner teacher identity development: external mentors and the third space. *Research Papers in Education* 31(2): 133-158.

Merton, R K (1968) *Social theory and social structure*. London: Simon and Schuster.

Michie, S, van Stralen, M M, & West, R (2011) The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science* 6(1): 42.

Muijs, D, & Reynolds, D (2001) Being or doing: The role of teacher behaviors and beliefs in school and teacher effectiveness in mathematics, a SEM analysis. In *annual meeting of the American Educational Research Association, Seattle, WA*

Nelson, J, & O'Beirne, C (2014) *Using evidence in the classroom: What works and why?* Slough: National Foundation for Educational Research.

Newman, F M, Rutter, R A, & Smith, M S (1989) Organizational factors that affect school sense of efficacy, community and expectations. *Sociology of Education* 62: 221–238.

Nutley, S M, Walter, I, & Davies, H T (2007) *Using evidence: How research can inform public services*. London: Policy press.

Pawson, R and Tilley, N (1997) *Realistic evaluation.* London: Sage.

Rogers, P J, & Weiss, C H (2007) Theory-based evaluation: Reflections ten years on: Theory-based evaluation: Past, present, and future. *New directions for evaluation* 114: 63-81.

Rogers, P J, Petrosino, A, Huebner, T A, & Hacsi, T A (2000) Program theory evaluation: Practice, promise, and problems. *New directions for evaluation* 87: 5-13.

Rogers, PJ (2008) Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14*(*1): pp.29-48.

Sharples, J, Albers, B, & Fraser, S (2018) *Putting evidence to work: a school's guide to implementation. Guidance Report.* London: Education Endowment Foundation.

Sharples, J, Webster, R, and Blatchford, P (2015) *Making the best use of Teaching Assistants: Guidance Report*. Available at: https://educationendowmentfoundation.org.uk/public/files/Publications/Campaigns/TA_Guidance_Report_MakingBestUseOfTeachingAssisstants-Printable.pdf

Sheffield Institute of Education (2017) *Integrating English: trial protocol (amended November 2017)*. Available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Integrating_English_Protocol_Amended_2017.11.07.pdf

Sin, C (2008) The role of intermediaries in getting evidence into policy and practice: some useful lessons from examining consultancy–client relationships. *Evidence & Policy* 4 (1): 85-103.

Stame, N (2004) Theory-based evaluation and types of complexity. *Evaluation* 10(1): 58-76.

Stame, N (2010) What doesn't work? Three failures, many answers. *Evaluation* 16(4): 371-387.

Taylor-Powell E, Jones A L & Henert E (2003) *Enhancing program performance with logic models.* Available at: https://fyi.uwex.edu/programdevelopment/files/2016/03/lmcourseall.pdf

Tilly, C (2005) *Identities, boundaries and social ties*. London: Routledge.

Timperley, H, Wilson, A, Barrar, H, & Fung, I (2007) *Teacher professional learning and development: Best evidence synthesis iteration.* Auckland, NZ: Ministry of Education.

Tschannen-Moran, M, & Hoy, A W (2001) Teacher efficacy: Capturing an elusive construct. *Teaching and teacher education* 17(7): 783-805.

Tschannen-Moran, M, Hoy, A W, & Hoy, W K (1998) Teacher efficacy: Its meaning and measure. *Review of educational research* 68(2): 202-248.

University of Wisconsin (undated) *Logic Models.* Available at: https://fyi.uwex.edu/programdevelopment/logic-models/

Walton, M, (2014) Applying complexity theory: a review to inform evaluation design. *Evaluation and program planning* 45: 119-126.

Weiss, C H (1997) Theory-based evaluation: Past, present, and future. *New directions for evaluation*, 76: 41-55.

Weiss, CH, (1995) Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. *New approaches to evaluating community initiatives: Concepts, methods, and contexts* 1: 65-92.

Wholey, J S (2003) Improving performance and accountability: Responding to emerging management challenges. *Evaluating social programs and problems: Visions for the new millennium*.pp. 43-61.

Yin, R K (2013) Validity and generalization in future case study evaluations. *Evaluation* 19(3): 321-332.

Yoon, K S, Duncan, T, Lee, S W Y, Scarloss, B, & Shapley, K L (2007) Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (NJ1)*.