

**Using logic models in mixed methods research: the example of the Integrating English randomised controlled trial**

COLDWELL, MIke, CULLINEY, Martin <<http://orcid.org/0000-0002-7329-8675>> and MOORE, Nicolas <<http://orcid.org/0000-0002-0829-5523>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/37029/>

---

This document is the Accepted Version [AM]

**Citation:**

COLDWELL, MIke, CULLINEY, Martin and MOORE, Nicolas (2026). Using logic models in mixed methods research: the example of the Integrating English randomised controlled trial. In: MORRISON, Keith and SEE, Beng Huat, (eds.) Handbook of Mixed Methods Research in Education. Elgar Handbooks In Education . Edward Elgar, 289-308. [Book Section]

---

**Copyright and re-use policy**

See <http://shura.shu.ac.uk/information.html>

## CHAPTER 17

### USING LOGIC MODELS IN MIXED METHODS RESEARCH: THE EXAMPLE OF THE INTEGRATING ENGLISH RANDOMISED CONTROLLED TRIAL

**Mike Coldwell, ORCID 0000-0002-7385-3077, Martin Culliney, ORCID. 0000-0002-7329-8675, and Nick Moore, ORCID 0000-0002-0829-5523**

#### **Abstract**

This chapter focuses on using logic models as a way of framing mixed methods approaches, which have come to prominence in randomised controlled trials (RCTs). The framework and rationale of logic models motivates a principled eclecticism in our approach to mixed methods research (MMR). It justifies the combination of data collection methods and analyses in measuring and testing the process and outcomes of large-scale research, including RCTs. We summarise an approach to logic models developed in earlier work (Coldwell & Maxwell, 2018) and exemplify its application to a specific RCT, Integrating English, highlighting prompts that can determine an appropriate mix of data collection methods. From these and other applications, we consider issues in the future and challenges involved in applying this approach in MMR. The case is made for MMR exposing the strengths and limitations of logic models and for providing rich, action-oriented data on intervention-based research such as RCTs.

**Keywords:** logic models; randomised controlled trials; linguistics; primary schools; context.

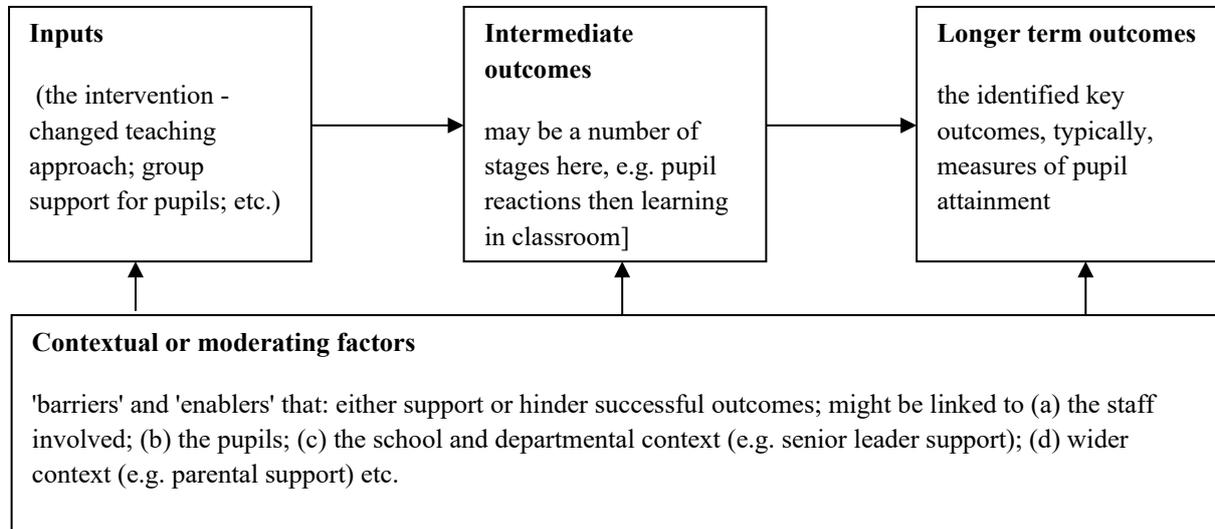
## **[a] 17.1 Introduction**

Logic models are widely used to frame mixed methods research and evaluation designs. They offer a principled, theory-driven approach to coherent mixed methods research, among other benefits (Cox, 2000; Rogers et al., 2000), and have come to prominence with the rise of randomised controlled trials (RCTs) to evaluate the impact of programmes and interventions across public policy fields, particularly in education (Education Endowment Foundation, 2024; What Works Network, 2024). A range of critiques have shaped developments in logic models, including Coldwell and Maxwell's (2018) 'evidence-informed logic model' approach. In this chapter, we discuss logic model designs, their critiques and redevelopment to inform mixed methods studies. The chapter presents the first detailed analysis of an RCT approach focussed on its logic model design that uses mixed methods research, reflecting on its use in the context of the critiques of such models. It concludes with a discussion of how such models can be used for future mixed methods studies, particularly – but not only – for RCT designs.

## **[a] 17.2 Logic Models**

Coldwell and Maxwell (2018, p. 268) describe logic models as 'as a type of causal model representing how an intervention can lead to outcomes' via a series of steps in a diagrammatic format, typically 'single path models between inputs/activities, outputs, and intermediate and longer term outcomes/impact'. Figure 17.1 shows a very simple frame for such a model. Their popularity relates to the clear articulation of the steps from inputs to outputs and outcomes, which can then be used by the evaluator to establish measures to test if the intervention or programme is working as intended. In addition, since they are typically developed collaboratively between the evaluator and the intervention designer/developer,

logic models build a shared understanding of how the intervention is expected to work, bringing out potentially unarticulated assumptions around variables, causes, and effects.



**Figure 17.1** *A generic education logic model frame*

**[Alt-text: This figure moves from inputs to intermediate outcomes and thence to longer terms outcomes. These are rooted in ‘barriers’ and ‘enablers’ that: either support or hinder successful outcomes; might be linked to the staff involved, the pupils, the school and departmental context, and the wider context.]**

Logic models have proved to be particularly important for RCTs. In education, Lortie-Forgues and Inglis (2019) found that only a minority of RCTs show a positive impact (including less than 20% of those funded by the Education Endowment Foundation, henceforth EEF, in England). It was recognised many years ago by researchers working within the theory-based evaluation tradition that while experimental designs such as RCTs ‘give good estimates of program impact, they have little to say about how or why the impacts occurred’ (Weiss, 1997, p. 502). For this reason, the EEF requires the development of a logic model at the beginning of each RCT evaluation, noting ‘[t]he logic model developed in

collaboration between the evaluation and delivery team should be used to support the design of the IPE [Implementation and Process Evaluation] research questions, and its integration with IE [Impact Evaluation]' (EEF, 2022, p. 3)<sup>1</sup>. This quotation also identifies the importance of logic models in mixed methods designs, with its reference to integrating Implementation and Process Evaluation (referred to throughout this chapter as IPE) with Impact Evaluation. While the Impact Evaluation estimates the impact of the intervention on a given outcome such as pupil attainment using quantitative measures, the IPE examines what successful implementation looks like, drawing on measures determined by the outcomes predicted in the logic model, and, as such, is agnostic with regard to quantitative and qualitative data and analysis. A typical combination of research methods in the IPE is exemplified by the Integrating English case study in this chapter.

However, since logic models have become widely used, a series of critiques have been presented. In earlier work (Coldwell & Maxwell, 2018), we identified four broad areas that are problematic for simple logic models of this kind, drawing on these critiques. Firstly, the simple model assuming a single pathway or chain cannot account for more complex designs, with multiple strands which may work together in parallel or phased ways; indeed, there may be more than one process taking place (Rogers, 2011). Secondly, logic models are often not designed to bring out the causal mechanism underlying the chains of inputs, outputs, and outcomes (Morris et al., 2016) – in the oft-cited phrase of Astbury and Leeuw (2010, p. 367), ‘mechanisms appear too frequently as unexplained ‘causal arrows’ that seem to flourish so well in the present climate of enthusiasm with visual logic models’. Further, even when a causal mechanism is identified, it is important for evaluators to consider that other causal mechanisms could be at work. Thirdly, whilst contextual factors may be considered (as in Figure 1 above), the complexity of the relationship between such factors and the programme is often underplayed. Coldwell (2019), for example, notes that contextual

factors: are dynamic not static; are often active ingredients in any change process; may influence inputs, outputs, and outcomes differently; and may act in a cyclical way acting as influences, outcomes and (over time) further inputs to the model<sup>2</sup>. Finally, logic models can struggle to deal with more complex programme designs, especially those that feature non-linearity, adaptation, uncertainty, and the emergence of outcomes as identified (Walton, 2014).

To try to deal with some of these issues, we developed what we refer to as an ‘evidence-based logic model’ design. Coldwell and Maxwell (2018) explain that this terminology is used to indicate two things : ‘Firstly, the model itself draws on research evidence presented above on the limitations of typical logic models’ summarised in the previous paragraph and, secondly, that the model itself ‘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’ (p. 279). The application of the evidence-based logic model approach is exemplified in the case study detailed below.

### **[b] *Logic models in mixed methods research***

As is clear from the contributions to this handbook, mixed methods designs serve a wide variety of purposes within a broad range of philosophical perspectives. Logic models within social evaluations are associated with positivist/empiricist or sometimes realist traditions. Certainly, experimental RCT designs such as the example presented here can be situated broadly in the positivist camp due to the assumption that the logic model represents relatively stable and predictable change processes observable via careful, systematic data gathering, which is related to the critiques and limitations of such models outlined in the previous section. Using Greene et al.’s (1989) five-fold categorisation of purposes of mixed methods

approaches, the role of logic models in RCTs accords most closely with the ‘complementarity’ purpose, with the IPE design used to elaborate and clarify the findings of the Impact Evaluation.

Discussion of the use of logic models to guide mixed methods designs is quite limited. Yin (2013), the eminent case study researcher, states that case study evaluations using mixed methods ‘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’ (p. 324). Most recently, Coldwell and Maxwell (2018) provide an extended discussion of how logic models can underpin mixed methods designs for an RCT pilot, and a local system change project evaluation. The key point of similarity here relates to the ability of the logic model to lay out features of the inputs, outputs, outcomes, contexts, and the interrelationships between them, which can then be used to help design appropriate methods to examine each.

However, as noted in the introduction, there has not previously been a sustained examination of logic models used to underpin RCT (plus IPE) designs, despite their increasing ubiquity. In an RCT context, the evidence-based logic model developed by Coldwell and Maxwell (2018) was built on learning from the case study discussed below: as these authors stated in 2018 (p. 279), ‘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'’. In the next section, we introduce this case study, which used data from pupil assessments, semi-structured interviews with teachers and school leaders, fieldwork observations in classrooms and training sessions, and documentation to measure and understand the impact of this programme on pupil attainment, and its effects on teachers and schools. The data were variously subject to

multilevel linear regression, corpus analysis, descriptive statistical measures and thematic analysis.

### **[a] 17.3 An Early Case Study: Integrating English**

#### **[b] *The programme and logic model***

The study took place in England from 2016 to 2018, funded by the EEF, The Bell Educational Trust and Philanthropy Unbound. The *Integrating English* programme was built around the LiLAC (Language in Learning Across the Curriculum) scheme (Custance, Dare, & Polias, 2012), which drew on systemic functional linguistics and genre pedagogy theory to improve pedagogy for subject-based literacy in schools (Rose & Martin, 2012). Integrating English aimed to improve linguistic proficiency and communicative ability for Years 5 and 6 students (age 9-11 years) and was expected to have the greatest impact on pupils with English as an Additional Language (EAL). The model is detailed in Culliney et al. (2019), but in essence it involved training teachers who were expected to develop, implement, and refine schemes of work, leading to changes in classroom teaching practice and, it was hoped, student performance.

The logic model (Figure 17.2) was developed through an iterative process involving the main funder (EEF), the delivery team at Enfield Council, and the independent evaluation team from Sheffield Hallam University, UK. In this model, rather than a single, simple pathway, there are two related pathways, one tracing teacher to school outcomes, and one from teacher to pupil outcomes. The potential causal mechanisms are not laid out fully, but are included here as *Enabling Characteristics*, as are a set of *Contextual Characteristics* that could influence the likelihood of the intervention being enacted (emboldened terms relate to key features in the logic model).

## Logic model for Integrating English, July 2016

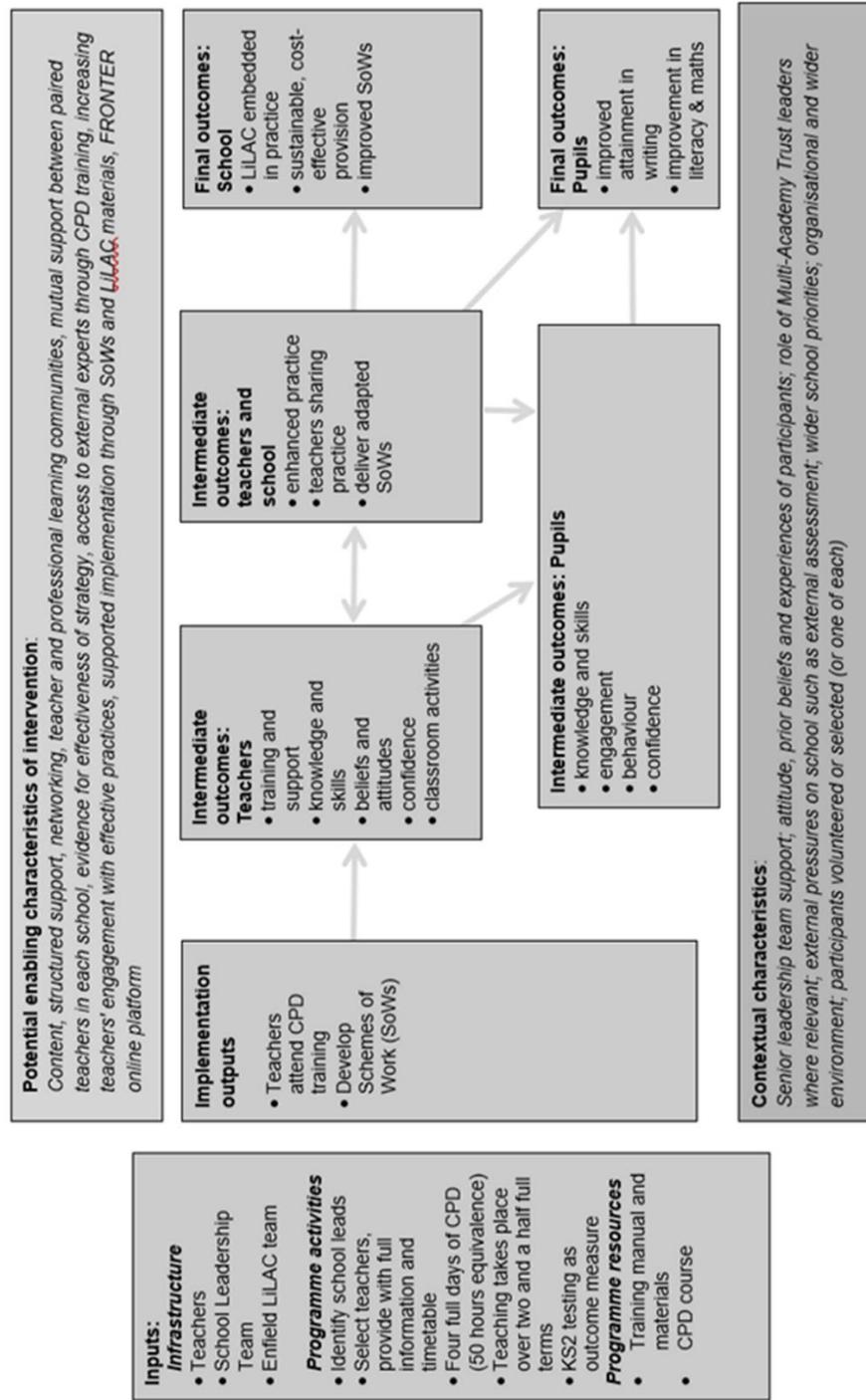


Figure 17.2 *Logic model for Integrating English*

**[Alt-text: This figure indicates initial inputs of infrastructure, programme activities, and programme resources. Then the model moves to implementation outputs (teacher development and schemes of work)), and then intermediate outcomes for teachers, schools, and pupils, leading to final outcomes for schools and pupils. The model takes significant account of multi-dimensional contextual characteristics.]**

**[b] *The trial design***

RCTs funded by EEF are classified according to the perceived stage of development of the intervention under evaluation. The Integrating English study was categorised as an efficacy trial, as the intervention had previously been piloted in a separate study conducted by the developer, involving ten schools. Efficacy trials test ‘interventions under ideal conditions and with developer involvement’ in contrast to effectiveness trials ‘that test a scalable model of the intervention in a greater number of schools’ (Dawson, Yeomans, & Brown, 2018, p. 293), aiming to be closer to ‘real-world conditions’ (Maxwell et al., 2021). Interventions showing promise at the efficacy stage may then be commissioned for an effectiveness trial, but both types of trial involve an Impact Evaluation and an IPE.

For the Integrating English study, the Impact Evaluation examined the effect of Integrating English on the attainment of all pupils, with further subgroup analysis on EAL pupils. Attainment, or *Final Outcomes for Pupils*, was measured through a writing assessment task administered by the evaluators, using a past paper from a recently discontinued statutory test of writing for pupils in the final year of primary (elementary) school. The IPE studied how the programme was implemented in schools and the experiences of teachers delivering the intervention, focusing on *Implementation Outputs* and *Intermediate Outcomes*.

The trial design was a two-arm, two-level clustered randomised controlled trial. Pupils were clustered into schools, which were randomly allocated to either (a) the intervention group, where schools delivered the programme over the study period, or (b) the control group, which continued teaching with a business-as-usual approach. Intervention schools selected the teachers to participate in the LiLAC training provided by the programme developer. These teachers then taught their pupils using the LiLAC approach throughout the study period. Schools allocated to the control group taught on a 'business-as-usual' basis during this time. The sample was well balanced in terms of key characteristics including geographical location, ratings awarded by the national schools' inspectorate (OFSTED), number of classes per school, pupil attainment, the proportion of pupils defined as disadvantaged, the proportion of pupils defined as EAL, and baseline test score (see Culliney et al., 2019).

Schools were recruited across England. These were divided into five geographical regions for the purpose of organising teacher training for schools randomly allocated to the intervention group. The trial had no teacher eligibility criteria. Schools were responsible for nominating the teachers to participate in the programme who would be teaching Y5 classes in 2016/2017 and Y6 in 2017/2018. The original intention was to recruit 100 schools, with a minimum of ten EAL pupils per school. However, the delivery team was only able to recruit 91 schools, despite relaxing the eligibility requirement to eight EAL pupils per school.

The baseline assessment, a commercially available language assessment administered in schools by the evaluation team, was completed by 4,763 pupils in October and November 2016. In June and July 2018, 3,607 pupils from 80 schools took part in the primary outcome assessment, an attrition rate of 24%. Secondary outcome data from statutory tests (on Reading and Grammar, Punctuation and Spelling) was obtained and analysed for the 90 schools remaining in the study.

## **[b] *The Implementation and Process Evaluation (IPE)***

The logic model for this project (see Figure 17.2), particularly the *Implementation Outputs* and *Intermediate Outcomes*, guided the IPE to select appropriate data on the programme activities (Befani, 2020), following guidelines designed by the funder (Humphrey et al., 2016). Budgetary and reporting constraints prevented the same approach being applied to *Final Outcomes for schools*.

A logic model is a visual representation of the theory of change underpinning research that assumes a cause-effect relationship, or multiple causal relationships (Jones et al., 2020). The logic model visually represents and makes explicit the key components of the relevant cause-effect relationships by itemising the dependent and independent variables assumed, by the theory of change, to effect change. As with any visual approach to meaning-making (Kress & van Leeuwen, 2020), logic models will highlight certain aspects of meaning, often through abstraction, while simultaneously obscuring other aspects. In the case of logic models, the processes of change are made less clear in this visual metaphor. As part of the Integrating English evaluation, the IPE investigated the change processes, hidden in the arrows in a logic model, which are critical to the success of the experiment conducted as an RCT, including some of the hypothesised potential *Enabling* and *Contextual Characteristics* of the project. Methodological choices for *Programme Activities*, *Implementation Outputs* and *Intermediate Outcomes* are discussed below.

To evaluate one *Programme Activity* and an *Implementation Output*, each of the five regional training centres that provided the programme CPD received a visit from the evaluation team member assigned to that region for one of the five CPD sessions. In this visit they were able to announce themselves and their intentions for the evaluation to the participants in that region irrespective of whether the school was visited for case studies.

Thus, the IPE team collected data on the complete training course and all participating teachers (minus the absentees) met a representative of the evaluation team.

The evaluation team focused on the quality and objectives of the CPD programme, guided by an observation scheme informed by ‘best practice’ in CPD (Maxwell et al., 2018). Evaluators were able to: monitor the adaptation of the programme from the Australian context (Custance, Dare, & Polias, 2012) to education settings in the UK; observe the techniques employed by trainers to engage workshop participants and monitor participants’ responses, both formally and informally; and to design follow-up questions about the CPD for case study interviews through feedback to the lead evaluator. In most cases, evaluators found that the CPD sessions reached high standards, were well-designed and well-received by attendees. Completion of the training programme, one of the measures of fidelity, was very high.

The *Implementation Output* of ‘Schemes of Work’ that showed incorporation of the Integrating English programme into teaching schedules were collected during and at the end of the trial and were, along with school-wide action plans, significant documentary evidence of how consistently schools had shown fidelity to the CPD programme (see Culliney et al., 2019; Moore, Coldwell, & Perry, 2021). While there was a rudimentary scheme to indicate the efficacy of these documents in the school context, submission was sufficient to indicate fidelity. Fidelity, however, was not high according to this measure – only one third of participating schools submitted both sets of documents. Consequently, the overall efficacy of Schemes of Work and Action Plans did not score highly.

Evaluation of the *Intermediate Outcomes* of the training programme was largely achieved through case studies (Hamilton & Corbett-Whittier, 2012). All intervention schools were invited to self-select as case study schools. Each case study consisted of an observation of two or more classes and interviews with at least two teachers and one school leader

directly involved in the trial. Visits were scheduled at the convenience of the school and evaluators requested to see classes that implemented what was learned from the Integrating English CPD programme.

A wide range of the issues covered by *Contextual Characteristics*, *Potential Enabling Characteristics* (including causal mechanisms) and *Intermediate Outcomes* were addressed in interviews and observed in classrooms during visits to case study schools. Classroom Observations (Hardman & Hardman, 2017) were designed to collect evidence showing that the CPD programme had impacted teaching practice. The observation proforma included a particular focus on the classroom setting, with ‘thick description’ (Kaplan-Weinger & Ullman, 2014) including the collection of artefacts or photographs (of student work, classroom displays, or teacher planning materials). It also featured specific items on how the intervention can change teaching practices, such as ‘Teacher uses ‘genre pedagogy’ technique or stage in class’.

Observers were primed to record evidence of the key characteristics of the Integrating English programme. Data enabled a comparison of the impact reported by case study teachers with the practices that could be observed independently. Results from the IPE case studies, including a deductive Framework Analysis (Saldana, 2011; Smith & Davies, 2010), showed that despite a very positive response by teachers in interviews, observations produced little evidence of fundamental change in pedagogy (Culliney et al., 2019). Corpus linguistic analysis (Pérez-Paredes, 2020) of interview data produced a picture of the CPD programme being very successful in some areas, such as techniques for working with students in their class who used English as an Additional Language (EAL), but less impactful on practice in other areas, such as the key change of adopting the teaching-learning cycle proposed for genre-based pedagogy (Rose & Martin, 2012).

A post-intervention survey of teachers (Lodico, Spaulding, & Voegtler, 2006; Cohen, Manion, & Morrison, 2018) was designed to measure *Intermediate pupil outcomes* (*knowledge and skills, engagement, behaviour, confidence*) from the teachers' perspective. The survey aimed to capture change mechanisms that would indicate how well the intervention was achieving the final set of intermediate outcomes culminating in the subsequent Impact Evaluation results. As such, this can be seen as a bridge between the Impact Evaluation and the IPE. Intermediate outcomes measured in this way are useful both in their own right and as indicators of whether the programme is being implemented according to the logic model. A survey also provides a broader range of respondents than those sampled in the case studies for the IPE, offering greater representation.

The questionnaire was distributed electronically during the same period as the post-intervention assessments for the Impact Evaluation. At this time, the priority of the evaluation team was ensuring that schools took part in the assessments, as these would provide data for the primary outcome measure. This required a lot of time and effort in reaching the schools, where in some cases the main contact had moved on during the study period, which spanned two school years. Subsequently, the response rate for the teacher survey was so low that it was not possible to make secure judgments about the results, and it was decided to omit them from the final evaluation report.

#### **[b] *The Impact Evaluation (IE)***

*Final pupil outcomes* for the Integrating English logic model (Figure 17.2) were all centred on pupil attainment. It was agreed by all parties involved in the study - the evaluators, the trial funder, and the intervention delivery team - that pupil writing skills should be the primary outcome measure for the Impact Evaluation given the substance of the intervention. Selecting appropriate, valid outcomes is often a challenge for RCTs (Edovald & Nevill,

2021), and in this case the challenge was made greater after policy changes. This section reviews what happened when the logic model was applied to the Impact Evaluation of the Integrating English programme.

Stakeholders in the evaluation agreed to use the Key Stage 2 (KS2) writing test to measure programme impact. KS2 in England and Wales refers to pupils in Years 3 to 6, aged 7-11, with national testing in Year 6. The National Pupil Database (NPD) contains results from all statutory tests undertaken by pupils at schools in England, as well as relevant demographic data, and is the official source of individual level attainment data. Researchers must apply directly to the Department for Education, the government body responsible for schools in England, for extracts from the dataset. Using administrative data which is already collected for non-research purposes cuts costs, reduces the burden on participants and is less vulnerable to attrition (Dawson, Yeomans, & Brown, 2018). However, a policy change during the study period saw the KS2 writing test become teacher assessed. This meant it was not marked externally like other subjects such as maths or reading, with results no longer being made available through the NPD. Therefore, an alternative primary outcome measure was required.

The primary outcome measure, writing, was assessed through a test that was not in the initial evaluation design or the Memorandum of Understanding agreed with participating schools. In this instance, the delivery team questioned the validity of a general writing test focussed on the subject of English, when the intervention aimed to improve the use of language, including writing, across the curriculum in different subjects. However, as developing a bespoke measure was beyond the scope of the study (Culliney et al., 2019), an existing assessment or scale had to be used. After reviewing potentially suitable measures, a past KS2 writing paper, previously used as a statutory test, was chosen as the assessment

most closely aligned to the trial learning aims. This was added to the requirements for participating schools after they had signed the Memorandum of Understanding.

Ultimately, 11 of the 91 schools recruited to the trial did not take part in the post-intervention assessment, reducing the sample size for the final analysis. It is not clear whether this was because they objected to the additional test or for other reasons. Had a primary outcome measure from an administrative data source been available, attrition would probably have been lower as it was for the secondary outcomes, which were drawn from administrative data (under 5% compared to 24%). However, changes to the assessment of statutory tests were beyond the control of the evaluation team.

The impact of the intervention was measured using a multilevel model with pupils clustered into schools. The analysis controlled for pupil fluency in English, school size as defined by the number of Y5 classes, geographical location, and baseline assessment scores at both pupil and school level. Effect sizes were calculated using Hedges'  $g$ ; for the primary outcome, a measure of writing ability, this was -0.05 standard deviations (Confidence Intervals -0.21; +0.12,  $p = 0.577$ ). For the secondary outcomes, Reading and Grammar, Punctuation and Spelling, the effect sizes were -0.06 (Confidence Intervals -0.25; +0.12,  $p = 0.517$ ) and 0.10 (Confidence Intervals -0.25; +0.12,  $p = 0.917$ ) respectively. These results led the evaluators to conclude that the intervention was not associated with any improvements in pupil attainment in English.

### **[b] *Lessons from the Integrating English logic model***

Collecting teacher level data on the intermediate outcomes as part of the IPE was intended to contextualise the findings from the Impact Evaluation and provide an insider perspective on the intervention, as is established as a rationale for mixed method designs (Gorard & Taylor, 2004). The logic model predicts that if one impact in the cause-effect chain fails, the final

outcome will fail. The IPE provided verifiable data, through multi-method research design, to suggest that, and potentially how and where, the chain of effects between professional development and pupil attainment did not connect (Culliney et al., 2019). Further investigation into how this causal chain was enacted would have been required to understand, for instance, the role of CPD in changing pedagogy within a data-driven accountability regime in marketised education (Moore, Coldwell, & Perry, 2021). Evaluating the impact of CPD at school and teacher level was recognised as insufficiently researched (Desimone, 2009) and, where standard tools exist (Allen et al., 2021), they were subject to the same practical issues as the survey designed for the Integrating English evaluation.

Had sufficient returns of the teacher survey produced reliable data on pupil intermediate outcomes, the funder would not expect this evaluation data to take priority over the attainment data representing the final pupil outcomes. In fact, despite considerable data highlighting the processes involved in implementing the programme, summaries and headlines for the study focused almost exclusively on the results of the Impact Evaluation. The logic model hypothesised that improvements in pupil knowledge and skills, engagement, behaviour, and confidence would feed into improvements in attainment outcomes, but there was no intention to gather data systematically on those pupil intermediate outcomes or explore relationships with the final pupil attainment outcomes beyond the teacher surveys.

The logic model and the evaluation design more generally aimed to assess the impact of the intervention through pupil level attainment and to enhance the understanding of these findings through a series of other research methods - interviews, observations, documentation and surveys - where data collection treated teachers as units of analysis. The different units of analysis made method integration difficult, particularly as the two-level design of the Impact Evaluation did not account for teachers or classes, clustering pupils within schools and effectively ignoring the teacher or class level in the Impact Evaluation.

This raises the question of why a teacher or class level was not included in the Impact Evaluation design when the logic model relied heavily on teacher level intermediate outcomes. This would have been complicated, as the delivery period spanned two school years and both Y5 and Y6 teachers were supposed to attend LiLAC training, prepare schemes of work, and teach pupils using the LiLAC approach. With the teacher or class level not accounted for in the Impact Evaluation, it was important that the IPE design provided the perspective of these participants who were central to the delivery of the intervention. While tighter method integration may have supported a more comprehensive examination of the steps comprising the logic model, a design prioritising different data sources and units of analysis arguably produced a more complete picture of the intervention and how its implementation was experienced by those at the chalkface.

In mixed method designs, method integration can be strengthened through using the same units of analysis (Woolley, 2009; Yin, 2006), so collecting pupil level data on pupil intermediate outcomes would have facilitated analysis of their relationship with the final pupil outcomes. Unfortunately, collecting pupil level data on intermediate outcomes was beyond the scope of the study. Subjecting pupils to an additional test, for instance, would have been difficult to organise, and hard to justify ethically given the extra burden for pupils and school staff of the primary outcome assessments. It is also unclear whether satisfactory measures of these intermediate outcomes existed in a format that would have been possible to use in this study. This does not amount to a flaw in the logic model or evaluation design as such, yet it illustrates the difficulties faced in any attempt to interrogate this important link in the causal chain.

The effect of compliance with the intervention on the trial outcome was calculated using the complier average causal effect (Bloom 1984; Hewitt, Torgerson, & Miles 2006). However, compliance was defined at school level, using data on whether teachers had

completed intervention activities (CPD), while the impact of the intervention was measured at pupil level. The intermediate outcomes in the logic model were not accounted for in the compliance analysis or the impact analysis more broadly. If this had been possible, the links between those two components of the logic model could have been explored, allowing for greater integration between the Impact Evaluation and the IPE.

One way in which this could have been achieved was through a three-level trial design, with pupils clustered into teachers and teachers clustered into schools. This approach is more common in studies focussing on secondary schools. At primary schools in England, it is less usual for pupils to be grouped by ability, or to attend classes taught by different teachers for different subjects, in contrast to secondary schools, where pupils are normally taught by subject specific teachers and setting by ability is more frequent, particularly in maths (Demack, 2019). Although the trial design could have monitored teacher compliance with the intervention and used that data alongside pupil outcomes to examine the effect of teacher engagement on attainment, this would have increased data burdens on schools. For example, as the delivery period spanned two school years, it would have been necessary for schools to supply a list of which pupils were taught by which teacher in both study years, doubling the number of data submissions required and increasing the risk of missing data. Evaluation design often involves these trade-offs where there is no perfect solution.

More broadly, as we go on to discuss at the end of this chapter, developing methods to plausibly link impact with process evaluation findings is an area of development in the wider evaluation field.

### **[b] *Broader reflections on the Integrating English logic model***

As noted above, an efficacy trial is focussed on measuring the impact of an intervention in optimal conditions. As such, resources are oriented toward measuring this impact instead of

examining the relationship between the intermediate outcomes, which are not necessarily quantified, and final outcomes, which by definition are always quantified. The compliance analysis, intended to examine whether adherence to the programme was associated with differential impact on the primary outcome measure, posited a causal connection between programme activities, variously specified as *inputs*, *implementation outputs*, and *final pupil outcomes*, with this relationship mediated by *intermediate outcomes* for both teachers (measured) and pupils (not measured).

The logic model was designed to help understand the trial data, but the RCT data was not used to question or refine the logic model. If the trial was run again, there would be the opportunity to revisit the logic model in the context of the findings, but for trials showing no impact a follow-up trial is rarely seen, so this avenue is seldom explored and the possibilities of learning from failure remain unfulfilled (an issue explored in more depth by Coldwell and Moore, 2024). Had the trial shown positive impact, it would have been assumed that the *inputs*, *implementation outputs* and *intermediate outcomes* all fed into the improvements in the final pupil outcomes, but links between the *intermediate outcomes* (at both pupil and teacher level) and *final pupil outcomes* were not robustly explored.

It may not be possible to examine the relationship between every stage in the logic model. Within the confines of a funded study with limited time and budget, it is almost inevitable that long-term outcomes such as future attainment are out of the scope for studies focussing on one cohort of students; this is less a limitation of the logic model than recognition that some results may take longer to produce as some effects are more long-term. One conclusion to draw here is that the holistic nature of the intervention is perhaps difficult to pin down in the quantifiable manner required in an RCT. EEF has acknowledged the challenges in scaling up projects, and this case may be an example of where ‘some

interventions or questions lend themselves more readily to RCTs than others' (Edovald & Nevill, 2021, p. 55).

#### **[a] 16.4 Further Developments in the Use of Logic Models**

Continuing to take key learning from the case study, here we discuss some other uses of logic models in mixed methods designs to demonstrate their versatility. Evidence-informed logic models have underpinned subsequent RCT studies which align more closely to the process developed by Coldwell and Maxwell (2018). For example, the evaluation of REACH Primary (Culliney et al., 2021), which focussed on a reading intervention in primary schools delivered by Teaching Assistants, utilised the logic model reproduced in Figure 17.3. Compared with the Integrating English model, the causal mechanisms are more fully explicated: the pathways relating to each outcome are defined and the role of causal factors is more fully developed. It is possible to see both REACH Primary and Integrating English as using different iterations of the evidence-informed model, both of which are developments from simple logic models, as articulated in Table 17.1.

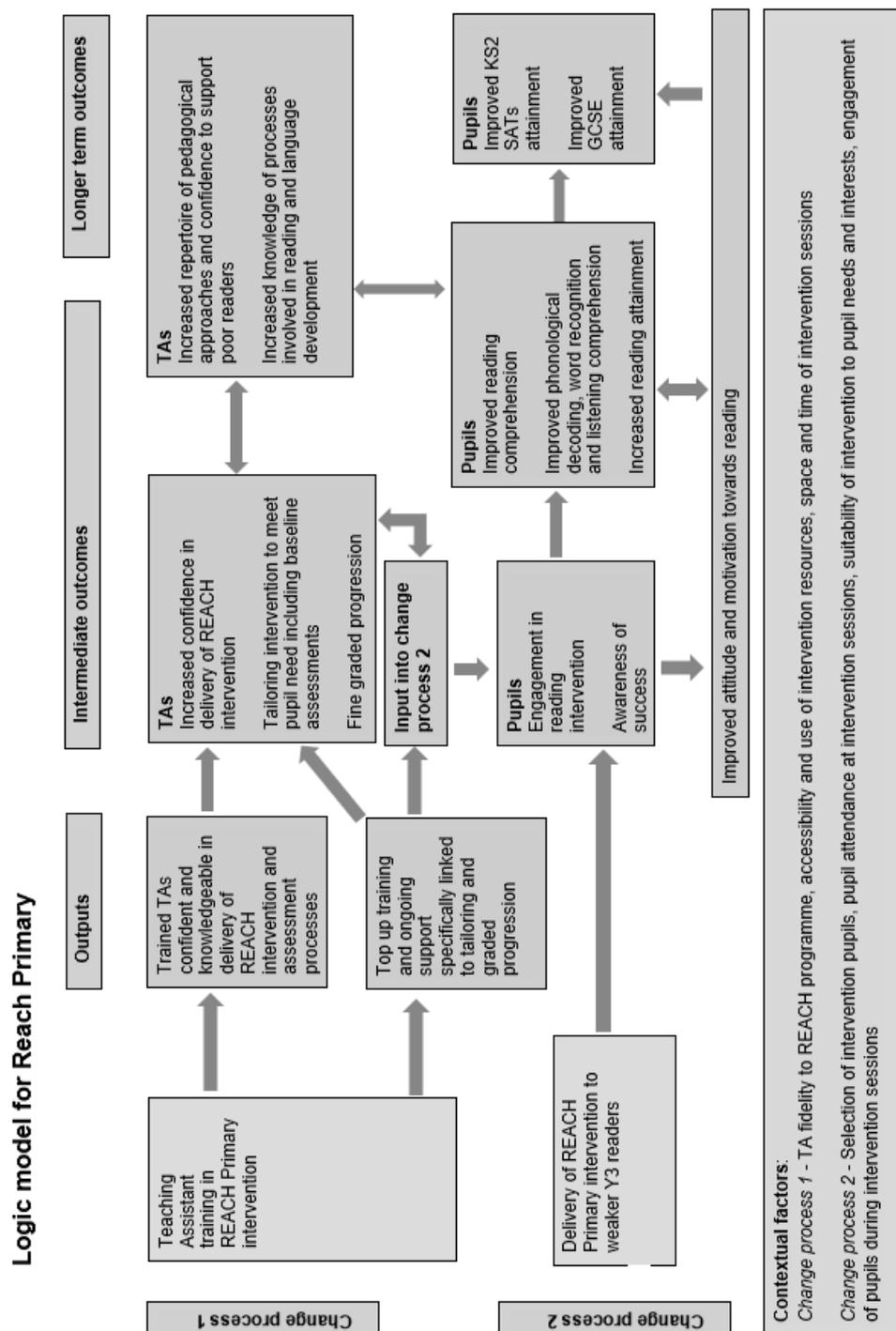


Figure 17.3 *Logic Model for REACH Primary*

[Alt-text: This figure operates in two change processes, and it moves from inputs (Teaching Assistants' development and interventions with weaker readers), to outputs (teacher performance and training), to intermediate outcomes and then longer-term

**outcomes for Teaching Assistants and pupils, including their attitudes and motivation towards reading.]**

**Table 17.1**

*Comparing simple and evidence-informed logic models*

<b>Simple model</b>	<b>Evidence-informed model</b>	<b>Example: Integrating English</b>	<b>Example: REACH Primary</b>
Single pathway from inputs via outputs to outcomes	Multiple pathways	Model includes two branches from teacher outcomes – one to school outcomes, the other to pupil outcomes, with some interactions noted	Two interrelated pathways are included with linkages made explicit
Causal mechanisms implicit	Causal mechanisms made explicit	Identified as enabling characteristics. But not fully explicated	Causal change processes laid out clearly and evidence underlying each is provided
Contextual factors ignored or their working not made clear	Contextual factors included and their operation clarified	Set of factors included, but not linked to different aspects of the model	Contextual factors included separately for each causal process

[Alt-text: The simple model of causality presents a linear input-output-outcome pathway, while the evidence-informed model includes multiple interconnected pathways with explicit causal mechanisms and contextual factors, for example: (a) the Integrating English example shows teacher outcomes branching to school/pupil outcomes; (b) the REACH Primary example links causal processes evidence and contextual factors.]

Versions of evidence-informed models have been used in other forms of mixed method evaluations. For example, a large and complex mixed method evaluation of the Teaching and Learning Innovation Fund, for the UK Government's Department for Education in England, that aimed to build the market for high quality professional development by commissioning a set of professional development sub-programmes delivered across England, included two versions of an overarching logic model (see Straw et al., 2022), in this case using the term 'theory of change'<sup>3</sup>. The evaluation design included a survey and a range of qualitative data, with a longer-term quasi-experimental impact design, quite different from the RCT evaluations above. Below this, separate models for each sub-programme were developed, closely modelled on the Coldwell and Maxwell (2018) approach. Figure 17.4 provides an example from the Geographical Association; it is recognisably similar in that it delineates two related change processes. These related, on the one hand, to teacher and school outcomes, (including pedagogic content knowledge, abbreviated to PCK) which included a focus on teacher retention, and, on the other, to changes in teachers' pedagogical approaches that were expected to lead on to pupil outcomes. As this was a less tightly defined programme, compared with RCT interventions, there was explicit mention made of potential complexity, as well as contextual factors. For the full evaluation report, see Wolstenholme et al. (2022).

TLIF GA/ASE: Evidence-informed logic model frame

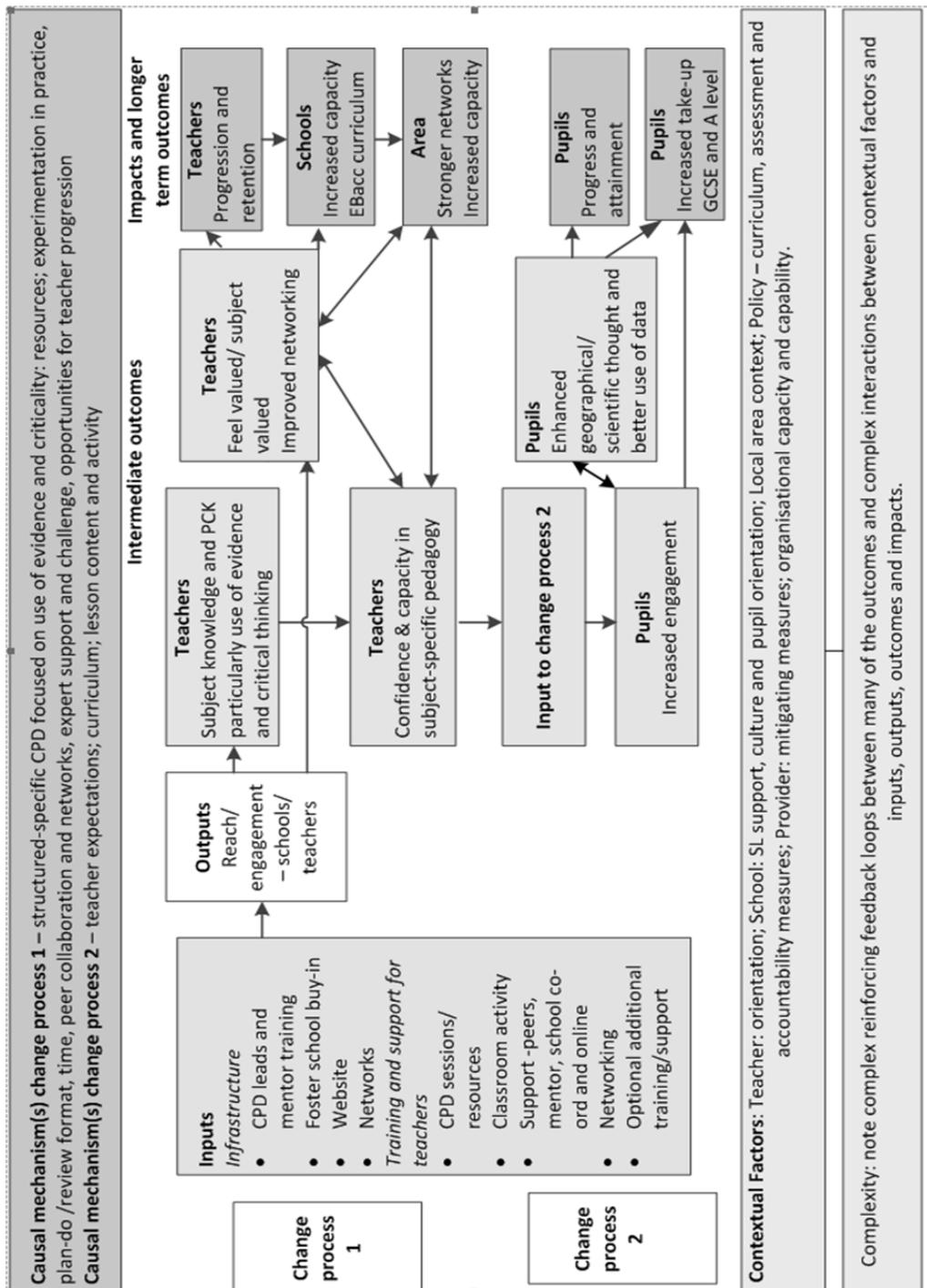


Figure 17.4 Logic model for Geographical Association evaluation

[Alt-text: This figure operates in two change processes, and it moves from inputs (infrastructure and training and support for teachers), to outputs, to intermediate outcomes for teachers, and pupils, and then impact and longer-term outcomes for

**teachers, schools, areas (networks), pupils, and including increased take-up of public examinations in Geography.]**

More broadly, readers may find it helpful to refer to the broad frame presented by Maxwell and Coldwell (2018, p. 292), should they choose to design a similar logic model to guide multi-method research or one of the many evaluation frameworks that rely on similar evidence-based logic models, some of which report advances in healthcare research (Jones et al., 2020; Mills, Lawton, & Sheard, 2019).

### **[a] 16.5 Challenges in Doing Mixed Method Research with Logic Models**

Our experience of using a logic model approach to inform the mixed method design of an evaluation highlights a set of challenges to be considered by researchers. Chief among these is the need to clarify and gather data on the causal mechanisms linking intermediate and longer-term outcomes. In the design phase, it is also important to ascertain the availability of appropriate data and the feasibility of collecting primary data, where required. This is particularly important when the logic model requires the involvement of different participant groups, when collecting data from them could be subject to different challenges. In this study, pupil assessment data became more difficult to gather following changes to the statutory testing regime, as described above. This led to increased demands on teachers who were required to administer an additional assessment, making it more of a challenge for them to submit survey responses. Mixed method researchers often aim for method integration and logic models can support this, but when the separate strands are interdependent, challenges with one data source or participant group can have broader implications.

The challenges encountered in this study will be familiar to all social science researchers using multi-methods designs. Here we highlighted the benefits of logic models to multi-methods research as a co-constructed, typically visual, representation of the hypothesised causes, causal mechanisms, effects, and outcomes to be monitored and measured, we proposed how we can research and reflect on logic models to improve their applicability, and we noted their limitations. We recognise that research studies operate within the real world and constantly make compromises between the best sources of data, resources and ethics (Befani, 2020). Additional theory-based approaches such as the use of Qualitative Case Analysis (see Cooper et al., 2012) or contribution analysis may be helpful to better predict and monitor the interaction of diverse sources and types of data (CECAN, 2023; HM Treasury, 2020a, pp. 43-45; HM Treasury, 2020b), a key deficiency of this approach to mixed method research, based on our reflections of evaluation projects.

As well as these challenges, it is perhaps crucial to acknowledge that for some programmes, whilst a mixed method design is appropriate, the logic model may not be. Logic models are not a one-size-fits-all solution to evaluations or to mixed method research design. It is unlikely that they can model with much fidelity the influence of all variables operating within an educational context or to manage complexity in interventions. Complexity may come in the form of multi-point chains of cause-and-effect relations (Rogers, 2011), such as the Integrating English evaluation (CPD changes teacher practice which changes student knowledge and behaviour which improves skills which is measured in test results). Complex interventions that may also trouble logic models are those that are implemented dynamically, with cycles of change dependent on research findings from a previous cycle, generating a learning spiral of innovation, research, and change. Mixed method designs that focus on complex system or organisational change programmes may be better suited to alternative models such as Connell and Kubisch's (1998) Theory of Change approach, deriving from

Complex Community Initiative evaluations – as Coldwell, Pearson and Wilson (2022) note ‘the ToC is focused on a system or organisational change process (such as a comprehensive community initiative). In contrast, a logic-model based approach is more suitable for a closed, programme-focused initiative’ (p. 451).

As will have become clear in this chapter, when it comes to mixed method evaluation design, logic models are not the only ‘game in town’. There are alternative approaches to laying out the expected steps and relationships between inputs, outputs and outcomes, and we warn how an RCT lens can produce a culture of educational research that limits definitions and perspectives (Burnett & Coldwell, 2021). Nevertheless, logic models are widely used by mixed method researchers for many reasons: they provide visual clarity; they can be co-created by researchers and other stakeholders; they can be informed by various forms of evidence; and they provide an agreed starting point to decide on the forms of data that need to be gathered. Because they are – to some extent at least – paradigm-neutral, there is no prior expectation of what data forms need to be used, with administrative, documentary, survey, interview, focus group, demographic, official, census and attainment data all in scope, where ethical and feasible. The key issue is that logic models can guide mixed method researchers to gather data to examine to what extent the anticipated pathways played out and the expected outcomes took place. Crucially, all data sources are plausibly connected via a theorised mechanism of change. While RCTs have some expectation of largely numerical data forms to be used in Impact Evaluation, a range of other data types, as discussed in the case study underpinning this chapter, can be used as part of the IPE to create a coherent design.

We offer these caveats around using logic models not to dissuade researchers from considering their use, but to critically present their value as tools as objectively as our experience allows. What the logic model displayed most clearly to the evaluation team were critical points in the intervention where subsequent changes were most likely to affect

outcomes. The mixed method approach exposed these very clearly. Guided by the logic model, the process evaluation produced data that questioned whether the intervention significantly impacted classroom pedagogy. The theory of change visualised by the logic model predicts that if this outcome is not achieved, any final impact is not the result of the intervention. Lack of impact is therefore probably not because students' test performance did not go up, or because students' knowledge and behaviour did not improve, but because teachers' pedagogy did not show significant impact from the CPD. With this knowledge, the intervention can be redesigned, taking on board lessons learned from this iteration (Coldwell & Moore, 2024): that a longitudinal approach to CPD is more likely to impact pedagogy, and that final impact measures need to be reconsidered, for instance.

The experience of conducting this evaluation also demonstrates the strengths of multi-method approaches. While certain elements of the design were not fully realised, the breadth of data sources, the diversity of participants and the longitudinal focus all ensured that sufficient data was collected to make a robust evaluation possible, offering a multifaceted perspective on the intervention being studied. Using a logic model approach laid the foundation for this evaluation, with defined programme inputs and outputs, clearly delineated causal pathways, and distinct outcomes for participants.

The use of the logic model in the Integrating English evaluation provided a valuable tool to create an integrated mixed method design, as well as highlighting improvements to be made to the model for future designs and research practice. In particular, the pathways laid out from inputs to outcomes allowed the evaluation team to focus the IPE data collection using optimal methods for the nature of data required to evidence the variables enabling change. Improved logic models were more effective in guiding multi-method research design for evaluation, with final impacts often measured without an RCT, and they achieved greater integration between units of analysis as well as a better combination of data types.

## **[a] 16.6 Conclusion**

The logic model in the Integrating English evaluation proved a useful tool to create an integrated mixed method design and to enable reflection for future developments. In particular, the pathways laid out from inputs to outcomes guided the evaluation team to collect data for the IPE that provided evidence of impact and change; and engaging both the funder and the delivery team in constructing the model ensured that all stakeholders were involved and in agreement about the details of the evaluation. The challenges encountered around the primary outcome measure enabled the team to draw out learning for future studies, and practical challenges – such as engagement with the survey – also provided important lessons.

An ideal outcome of such reflections would be further refinement of the logic model and evaluation design for future adaptations of the intervention; a key insight here is that the logic model interprets the statistical outcome measures against the implementation and process findings to better understand the results and so judgements around fidelity, efficacy, retrievals, further developments and so on should be considered holistically rather than affording priority to one set of measures (see Coldwell and Moore, 2024 for further discussion). As indicated in the Further Developments section, logic models can continue to be improved for RCT and other mixed method designs. Logic models are rarely subjected to such scrutiny as provided in this chapter, so we encourage mixed method researchers, evaluators and other stakeholders to continue to apply reflection and adaptation to their models.

Mixed methods are invariably the logical result when a logic model is used to determine the type of data collection and analysis for a research project or evaluation in the

complex context of education. We describe logic models as a paradigm-neutral, mixed method approach to research and evaluation design. Logic models select appropriate, feasible and ethical methods of data collection to measure each output and outcome and to monitor the mechanisms of change. The logic model generates research questions and so these results are then interpreted against the hypotheses.

In our experience as evaluators, we do not expect to accurately assess the ability to enact change in the classroom with the use of numerical indicators, such as test scores, alone. In optimal circumstances, the most reliable results from an RCT can only tell us *if* a change took place. When we need to observe and measure change mechanisms, such as the embodied processes of learning and of developing pedagogy, to better understand *how*, *why* or *in what conditions* change takes place, it is only logical to research using mixed methods.

#### **[a] Notes**

1. In EEF trials, Impact Evaluation estimates the impact of the intervention on a given outcome such as pupil attainment, while the Implementation and Process Evaluation examines what successful implementation looks like and how it could be scaled-up.
2. A good illustration is the influence of school leadership support for an intervention: this is an influence on the success of the initiative, but it may change (if senior leaders change) and can itself be influenced; if there is apparent success, leaders may switch from being sceptical to being more supportive [and vice versa]; and this support is likely to be particularly important in the early stages of the project.
3. There is not space to discuss in detail here, but – as with the term ‘logic model’ - ‘theory of change’ is used in a variety of different ways in evaluation designs, sometimes as a virtual synonym for logic model, as was the case here, and sometimes

rather differently in relation to types of non-experimental evaluation research designs particularly those focused on local system and organisational change (see Coldwell, Pearson, and Wilson, 2022, p. 450-452 for an overview of this second meaning in theory of change).

## [a] References

- Allen, L. M., Palermo, C., Armstrong, E., & Hay, M. (2021). Measuring impacts of continuing professional development (CPD): The development of the CPD impacts survey (CPDIS). *Medical Teacher*, 43(6), 677–685.  
<https://doi.org/10.1080/0142159X.2021.1887834>
- Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31(3), 363–381.  
<https://doi.org/10.1177/1098214010371972>
- Befani, B. (2020). *Choosing appropriate evaluation methods: A tool for assessment & selection*. Version 2. CECAN. [https://www.cecan.ac.uk/wp-content/uploads/2020/11/Final\\_Choosing-Appropriate-Evaluation-Methods-1.pdf](https://www.cecan.ac.uk/wp-content/uploads/2020/11/Final_Choosing-Appropriate-Evaluation-Methods-1.pdf)
- Bloom, H.S. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review*, 8(2) 225-246
- Burnett, C., & Coldwell, M. (2021). Randomised controlled trials and the interventionisation of education, *Oxford Review of Education*, 47(4), 423-438.  
<https://doi.org/10.1080/03054985.2020.1856060>
- CECAN (2023). *The Centre for the Evaluation of Complexity Across the Nexus*. CECAN  
<https://www.cecan.ac.uk/>

- Cohen, L., Manion, L., & Morrison, K. R. B. (2018). *Research Methods in Education* (eighth edition). Routledge.
- Coldwell, M. (2019). Reconsidering context: Six underlying features of context to improve learning from evaluation. *Evaluation*, 25(1), 99–117.  
<https://doi.org/10.1177/1356389018803234>
- Coldwell, M., & Maxwell, B. (2018). Using evidence-informed logic models to bridge methods in educational evaluation. *Review of Education*, 6(3), 267–300.  
<https://doi.org/10.1002/rev3.3151>
- Coldwell, M., & Moore, N. (2024). Learning from failure: A context-informed perspective on RCTs. *British Educational Research Journal*, 50(3), 1043–1063.  
<https://doi.org/10.1002/berj.3960>
- Coldwell, M., Pearson, S., & Wilson, I. (2022). Evaluating local system change using a comparative maturity matrix. *Evaluation*, 28(4), 446–465.  
<https://doi.org/10.1177/13563890221126744>
- 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, & J. Connell (Eds.) *New approaches to evaluating community initiatives: theory, measurement, and analysis, Vol.2.* (pp. 15-44)  
Washington, DC: The Aspen Institute.
- Cooper, B., Glaesser, J., Gomm, R., & Hammersley, M. (2012). *Challenging the qualitative-quantitative divide: Explorations in case-focused causal analysis.* Bloomsbury.
- 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. <https://doi.org/10.1177/030802260006300305>

- Culliney, M., Daniels, K., Booth, J., Coldwell, M., & Demack, S. (2021). *REACH Primary - evaluation report*. Education Endowment Foundation.  
[https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/REACH\\_report\\_v5-final.pdf](https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/REACH_report_v5-final.pdf)
- Culliney, M., Moore, N., Demack, S. & Coldwell, M. (2019). *Integrating English - Evaluation report*. Education Endowment Foundation.  
[https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Integrating\\_English.pdf](https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Integrating_English.pdf)
- Custance, B., Dare, B., & Polias, J. (2012). *Teaching ESL students in mainstream classrooms: Language in learning across the curriculum*. Department of Education and Child Development (DECD).
- Dawson, A., Yeomans, E., & Brown, E. R. (2018). Methodological challenges in education RCTs: Reflections from England's Education Endowment Foundation. *Educational Research*, 60(3), 292–310. <https://doi.org/10.1080/00131881.2018.1500079>
- Demack, S. (2019). *Does the classroom level matter in the design of educational trials? A theoretical and empirical review*. Education Endowment Foundation.  
[https://educationendowmentfoundation.org.uk/public/files/Publications/Does\\_the\\_classroom\\_level\\_matter.pdf](https://educationendowmentfoundation.org.uk/public/files/Publications/Does_the_classroom_level_matter.pdf)
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- Edoald, T., & Nevill, C. (2021). Working out what works: The case of the Education Endowment Foundation in England. *ECNU Review of Education*, 4(1), 46–64.  
<https://doi.org/10.1177/2096531120913039>

Education Endowment Foundation (2022). *Implementation and process evaluation guidance for EEF evaluations. Updated August 2022*. Education Endowment Foundation.

<https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-IPE-Guidance-August-2022.pdf>

Education Endowment Foundation (2024). *Evaluation design. Resources to support evaluation design and analysis*. Education Endowment Foundation

<https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design>

Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research*. Open University Press.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274. <https://doi.org/10.3102/01623737011003255>

Hamilton, L., & Corbett-Whittier, C. (2013). *Using case study in education research*. Sage.

Hardman, F., & Hardman, J. (2017). Observing and recording classroom processes. In D. Wise, N. Selwyn, E. Smith & L. E. Suter (Eds.) *The BERA/SAGE Handbook of Educational Research*. Sage.

Hewitt, C.E., Torgerson, D.J., & Miles, J.N.V. (2006) Is there another way to take account of noncompliance in randomized controlled trials? *Canadian Medical Association Journal*, 175(4), 347-348

HM Government (2024) *What Works Network*. HM Government.

<https://www.gov.uk/guidance/what-works-network>

HM Treasury (2020a). *Magenta Book. Central government guidance on evaluation*. HM Treasury.

[https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT\\_Magenta\\_Book.pdf](https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT_Magenta_Book.pdf)

HM Treasury (2020b). *Magenta Book Annex A. Analytical methods for use within an evaluation*. HM Treasury.

[https://assets.publishing.service.gov.uk/media/5e96c41a86650c2dd9e792ea/Magenta\\_Book\\_Annex\\_A\\_Analytical\\_methods\\_for\\_use\\_within\\_an\\_evaluation.pdf](https://assets.publishing.service.gov.uk/media/5e96c41a86650c2dd9e792ea/Magenta_Book_Annex_A_Analytical_methods_for_use_within_an_evaluation.pdf)

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016).

*Implementation and process evaluation (IPE) for interventions in educational settings: A synthesis of the literature*. Education Endowment Foundation.

[https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/IPE\\_Review\\_Final.pdf](https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/IPE_Review_Final.pdf)

Jones, N. D., Azzam, T., Wanzer, D. L., Skousen, D., Knight, C., & Sabarre, N. (2020).

Enhancing the effectiveness of logic models. *The American Journal of Evaluation*, 41(3), 452–470. <https://doi.org/10.1177/1098214018824417>

Kaplan-Weinger, J., & Ullman, C. (2014). *Methods for the ethnography of communication:*

*Language in use in schools and communities*. Routledge.

Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (4th Edition). Routledge.

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2006). *Methods in educational research:*

*From theory to practice*. Jossey-Bass.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often

uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166.

<https://doi.org/10.3102/0013189X19832850>

Maxwell, B., Clague, L., Byrne, E., Culliney, M., Coldwell, M., Hobson, A., & Glentworth,

A. (2018) *Retain: CPD for early career teachers of KS1 – Pilot report and executive*

*summary*. Education Endowment Foundation.

<https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/retain>

Maxwell, B., Stiell, B., Stevens, A., Demack, S., Coldwell, M., Wolstenholme, C., Reaney-Wood, S., & Lortie-Forgues, H. (2021). *Review: Scale-up of EEF efficacy trials to effectiveness trial*. Education Endowment Foundation.

[https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-syntheses/qualitative\\_analysis\\_of\\_factors\\_influencing\\_scale\\_up\\_from\\_efficiency\\_to\\_effectiveness\\_trials.pdf](https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-syntheses/qualitative_analysis_of_factors_influencing_scale_up_from_efficiency_to_effectiveness_trials.pdf)

Mills, T., Lawton, R., & Sheard, L. (2019). Advancing complexity science in healthcare research: The logic of logic models. *BMC Medical Research Methodology*, *19*(1), 55–55. <https://doi.org/10.1186/s12874-019-0701-4>

Moore, N., Coldwell, M., & Perry, E. (2021). Exploring the role of curriculum materials in teacher professional development. *Professional Development in Education*, *47*(2–3), 331–347. <https://doi.org/10.1080/19415257.2021.1879230>

Morris, S. P., Edovald, T., Lloyd, C., & Kiss, Z. (2016). The importance of specifying and studying causal mechanisms in school-based randomised controlled trials: Lessons from two studies of cross-age peer tutoring. *Educational Research and Evaluation*, *22*(7–8), 422–439. <https://doi.org/10.1080/13803611.2016.1259113>

Pérez-Paredes, P. (2020). *Corpus linguistics for education: A guide for research*. Routledge.

Rogers, P. J. (2011). Implications of complicated and complex characteristics for key tasks in evaluation. In K. Forss, M. Marra, & R. Schwartz (Eds.). *Evaluating the complex: Attribution, contribution and beyond* (pp. 33-52). Routledge.

- Rogers, P. J., Petrosino, A., Huebner, T. A., & Hacsí, T. A. (2000). Program theory evaluation: Practice, promise, and problems. *New Directions for Evaluation*, 87, 5–13. <https://doi.org/10.1002/ev.1177>
- Rose, D., & Martin, J. (2012). *Learning to write, reading to learn*. Equinox.
- Saldana, J. (2011). *Fundamentals of qualitative research*. Oxford University Press.
- Smith, K., & Davies, J. (2010). Qualitative data analysis. In L. Dahlberg & C. McCaig (Eds.), *Practical research and evaluation* (pp. 145-158). Sage. <https://doi.org/10.4135/9781446268346.n10>
- Straw, S., Nelson, J., Maxwell, M., & Coldwell, M. (2022). *Teaching and Leadership Innovation Fund (TLIF): Fund-level evaluation final report*. Department for Education. [https://assets.publishing.service.gov.uk/media/631f678ee90e077dc154d3a3/Teaching\\_and\\_Leadership\\_Innovation\\_Fund\\_\\_TLIF\\_\\_Fund-level\\_evaluation.pdf](https://assets.publishing.service.gov.uk/media/631f678ee90e077dc154d3a3/Teaching_and_Leadership_Innovation_Fund__TLIF__Fund-level_evaluation.pdf)
- Walton, M. (2014). Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning*, 45, 119–126. <https://doi.org/10.1016/j.evalprogplan.2014.04.002>
- Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review*, 21(4), 501–524. <https://doi.org/10.1177/0193841X9702100405>
- Wolstenholme, C., Reaney-Wood, S., Maxwell, B., Straw, S., McLean, D., Worth, J., & Staunton, R. (2022). *Geographical Association/Association for Science Education Critical Thinking for Achievement project. Final report*. Department for Education. [https://assets.publishing.service.gov.uk/media/631f4d848fa8f501ff9bea8d/TLIF\\_evaluation\\_Geographical\\_Association\\_Association\\_for\\_Science\\_Education\\_Critical\\_Thinking\\_for\\_Achievement\\_project.pdf](https://assets.publishing.service.gov.uk/media/631f4d848fa8f501ff9bea8d/TLIF_evaluation_Geographical_Association_Association_for_Science_Education_Critical_Thinking_for_Achievement_project.pdf)

Woolley, C. M. (2009). Meeting the mixed methods challenge of integration in a sociological study of structure and agency. *Journal of Mixed Methods Research*, 3(1), 7–25.

<https://doi.org/10.1177/1558689808325774>

Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools*, 13(1), 41-47.

Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*,

19(3), 321–332. <https://doi.org/10.1177/1356389013497081>