Guidance on Theory of Change and logic model development for EEF-funded evaluations

Education Endowment Foundation

September 2022, version 1

INTRODUCTION

The EEF aims to fund interventions that have a strong theory of change (ToC), that is, a strong rationale for how the programme resources and activities are expected to lead to the desired outcomes. The EEF has recently introduced a ToC development process at application stage, where delivery teams are supported to develop a programme ToC. The ToC (and a visual representation in the form of a logic model) are further refined and finalised during the evaluation set-up and protocol writing stage. It is important that evaluation designs of the EEF-funded programmes are guided by the programme's ToC. This applies to both the impact evaluation and the implementation and process evaluation (IPE).[1]

Intervention developers and/or delivery teams[2] and evaluation teams both play important roles in producing a ToC and logic model for the evaluation of the programme. Delivery teams understand best what their programme consists of and what it is aiming to achieve. Evaluators are experienced in ensuring that logic models are measurable. Close collaboration between the teams is, therefore, essential.

This paper

1) explains what a ToC and a logic model are and why they are important,
2) describes the process and shared responsibilities of the delivery and evaluation teams in developing these, and
3) defines the elements of the logic model and introduces a standard logic model diagram.

WHAT IS A THEORY OF CHANGE AND A LOGIC MODEL, AND WHY ARE THEY IMPORTANT?

A **theory of change** (ToC) is a summary of the core components of a programme, its desired outcomes, and the expected causal relationships between these. In other words, the 'story' of what the programme is, how and why it is expected to work, and what results it is intended to achieve.

By **logic model** we mean a visual representation of the ToC covering a programme's inputs, activities, outputs, outcomes and underlying causal mechanisms (drawing on Coldwell &

[1] Note that this guidance is written for pilot evaluations, efficacy trials, and effectiveness evaluations.
[2] Referred to as delivery team from this point forward.

Maxwell, 2018; Cooksy, Gill, & Kelly, 2001; Funnell & Rogers, 2011; Kaplan & Garrett, 2005; Knowlton & Phillips, 2012).

Terminology varies in the evaluation literature. Funnell and Rogers (2011, pp. 23–24) list 22 phrases[3] that are sometimes used interchangeably with 'logic model', including 'theory of change'. At the EEF, we tend to refer to the visual representation of a programme's logic as the 'logic model', while the term 'ToC' may capture broader aspects of the programme's 'story', including a narrative rationale of *why* and *how* the programme works, and its underlying causal and contextual assumptions. However, we do not draw a strict distinction between the terms, and they may be used interchangeably.

We expect the programme ToC and logic model to support every stage of the evaluation, from design to interpretation and reporting (Cooksy et al., 2001; Knowlton & Phillips, 2012; Markiewicz & Patrick, 2016). The research design, research methods and research questions should clearly reflect the programme ToC and logic model. Data collection and analysis should be planned so that components and causal assumptions included in the logic model can be tested, and the extent to which they have been found to be supported by evidence should be discussed in the report.

In the event of a null or negative result, evaluations designed in close alignment with the logic model may be able to distinguish between theory failure (the programme does not work as hypothesised to achieve the expected outcomes), implementation failure (the programme was not implemented as intended) and methodology failure[4] (inadequate evaluation methods were selected, or suitable evaluation methods were used inadequately) (Coldwell & Maxwell, 2018; Stame, 2010). In the event of a positive result, evaluations designed in close alignment with the logic model may be able to provide evidence on how and why and for whom the intervention worked.

Process for developing and capturing the Programme and its ToC

Delivery teams and evaluation teams are jointly responsible for i) developing a detailed description of the programme, ii) describing the theoretical and evidence-based rationale behind the programme including how and why it may be expected to work, and iii) capturing the programme's logic in a visual diagram (the logic model).

EEF-funded evaluations use the following process for development of the programme description, ToC, and logic model:

1. The delivery team develops an initial description of the programme and its ToC with support from the EEF.

- **Initial application**: In their application, delivery teams describe in narrative form how the programme resources and activities are expected to lead to change in the desired

[3] These include 'causal chain', 'impact pathway', 'intervention logic', 'intervention theory', 'logical framework (logframe)', 'outcomes line', 'program logic', 'program theory', 'results chain', 'theory of action', 'theory of change'.

[4] Methodology failure can also lead to false positive results (Type I errors), concluding that a programme works when in fact it does not or there is insufficient evidence to support the positive conclusion. Likewise, methodology failure can lead to Type II errors, concluding that a programme does not work when, in fact, it does.

outcomes and provide supporting evidence for the hypothesised links. Delivery teams may attach a visual diagram to their application if one exists.

- **ToC workshops:** Shortlisted delivery teams participate in a General (introductory) ToC workshop and a programme-specific ToC workshop over a period of approximately 4-5 weeks. The workshops support delivery teams to complete a logic model diagram using a standard EEF template (see Appendix 1: EEF Logic Model Template). Delivery teams also capture the causal and contextual assumptions underlying their programme in an assumptions log (see Appendix 2: Assumption Logs). Note that the ToC workshops cover inputs, activities, outputs, and outcomes as well as the causal links between them, but usually do not cover moderators and unintended consequences.

- **Evaluator commissioning:** After the EEF's Grants Committee has provided initial approval for the programme, evaluators are commissioned. Evaluators receive a detailed description of the programme, the logic model, and the causal and contextual assumptions logs to support them in proposing an evaluation design that is well aligned to the programme ToC.

2. The delivery team and evaluation team collaboratively refine the logic model.

- **IDEA workshop:** Once an evaluator is selected, delivery and evaluation teams attend a collaborative IDEA workshop, where the teams refine the version of the logic model that will form the basis for the evaluation.

- **Programme description and rationale:** During this phase, the evaluator also starts completing a detailed description of the programme and its rationale for inclusion in the protocol, drawing on the TIDieR framework (see Appendix 3: TIDieR Framework). This description is based on information provided by the delivery team, which may be complemented by evaluators consulting the literature to fully describe the rationale and prior evidence for the programme.

3. The evaluation team finalises the description of the programme as it will be delivered for the purpose of the evaluation, its rationale, and the logic model for inclusion in the published protocol.

- **Project set-up meetings:** After the IDEA workshop, the logic model is handed over to the evaluator and it is their responsibility to finalise it for inclusion in the protocol, ensuring that the delivery team and the EEF have had a chance to review any changes proposed after the IDEA workshop.

- **Publishing the protocol:** The programme description, rationale, and logic model for the evaluation are included in the evaluation protocol, which is published on the EEF website. We recommend that the causal and contextual assumption logs as prepared by the delivery team are included in the protocol as an appendix. There is no expectation for the evaluator to update these as the logs represent the delivery team's assessment of the assumptions. Delivery teams may wish to update the logs after the IDEA workshop if any discussions during the workshop have warranted this.

4. The evaluation team tests the logic model and updates it based on data collected through the evaluation.

- **Data analysis:** After data collection, the evaluator tests the logic model by conducting pre-specified quantitative and qualitative analyses to evaluate the programme implementation and theory. Where the data analysis reveals that some aspects of the logic model were not supported by the evidence gathered, the evaluator revises the logic model in line with the evidence gathered during this evaluation.

- **Reporting:** The evaluator includes both the initial logic model developed with the delivery team and (where applicable), a revised version based on evaluation data in the final evaluation report, describing the extent to which the model was supported by the evidence, and describing how and why any changes from the initial model were made.

The above process applies when a project is first funded by the EEF. Projects that are re-granted will already have a programme description and logic model from the first evaluation. For these projects, the above process is adapted flexibly, and IDEA workshops typically focus on updating the logic model where the intervention has changed or where further refinement or specificity is required. It is helpful to document changes from the version of the logic model that was previously evaluated.

ELEMENTS OF THE LOGIC MODEL

There are hundreds of ways of visually depicting a logic model. It is important for the EEF to have a (relatively) consistent model across our projects with consistent language as this i) enables us to conduct a fair assessment of the strength of the programme theory at application stage, ii) enables us to provide guidance and support to delivery teams before evaluators are appointed, and iii) facilitates synthesis and communication of findings across reports.

The EEF logic model template is provided in Appendix 1: EEF Logic Model Template. We acknowledge that the template will not be perfect for every project, and we will work with delivery teams and evaluation teams to adapt the template to the project as required.

Principles when capturing a programme ToC in a logic model:

- **Programme-level:** The logic model needs to represent the specific version of the programme for the specific target population that will be evaluated. We recommend that the logic model captures all 'key' inputs, activities, outputs, and outcomes of the intervention as delivered for the evaluation. However, any one evaluation may not be able to investigate all aspects of the logic model. Evaluators may wish to grey out / colour code elements of the logic model that will not be investigated in the evaluation.

- **Balance specificity and simplicity:** The ToC workshops support delivery teams to represent their ToCs simply in a logic model diagram without losing specificity in the language used. When trying to represent a complex intervention in a simplified form, it can be hard to leave out elements that seem relevant. However, the 'story' behind a programme is often quite simple. The aim is to design a logic model that unpacks the complexity of the programme but is easily comprehensible.

- **Stratification by type of participant:** Many programmes have activities that occur at different levels – for instance, teacher training takes place at the teacher level, whereas the resulting teaching activities occur at the pupil level. The standard EEF template has separate rows for activities at the teacher and pupil level. These levels can be adapted – for example, some interventions may have activities at the level of a parent, a school leader, or a master trainer.

We recommend the following elements are included in EEF logic models.

Elements of the logic model covered in the ToC workshops

Inputs and activities: Activities are the actions, processes, or events conducted as part of the programme (e.g., CPD sessions, webinars, individual learning time using an adaptive computer-based software, small group tutoring time). Inputs are things that are required to carry out the activities, such as materials (e.g., textbooks, facilitator guides, software) and people (e.g., master trainers).

Guiding questions:

- What are the key features of the inputs and activities needed for the intervention to work as intended?
- What features of the inputs and activities differentiate the programme from other similar programmes or from usual practice?
- How often and for how long will activities be delivered?

To balance simplicity and specificity, it is helpful to focus on the (hypothesised) **core components**[5] of the intervention, i.e., those features that are expected to influence the intervention's success if implemented successfully and without which the intervention would be expected to be less effective.

It is important to use specific language when describing inputs and activities. For example, for a computer-based intervention that teaches phonics to struggling readers, it may be particularly important to state in the logic model that the starting point is *adaptive (i.e., tailored to the pupil's current skill level)* and that the activities are *progressive (i.e., become increasingly difficult)*. Core features of well-known pedagogical practices (e.g., dialogic reading) can be described in narrative form alongside the logic model if it is not possible to fit them into the diagram.

Outputs: Outputs are the direct results of the inputs and activities. You might consider:

- What will teachers / pupils do (more of or less of) as a direct result of the programme inputs and activities?
- What will teachers / pupils be exposed to that they may otherwise not have been?

[5] Core components are also referred to as active ingredients, critical components, or behavioural kernels.

For example, as a result of early years practitioners modelling key concepts with children daily during play-based activities (pupil-level input), children have more opportunities to hear and explore the key concepts (pupil-level output).

Short-term and long-term outcomes: Outcomes are the changes in knowledge, skills, attitude and/or behaviour that are expected in the groups targeted by the inputs/activities and outputs.

Short-term outcomes (or proximal outcomes) happen before long-term outcomes and happen as a direct result of programme outputs. Long-term outcomes happen after the short-term outcomes, and often as a result of the short-term outcomes (or in part as a result of them).

It can be difficult to define what is meant by short- and long-term. For EEF-funded evaluations, it is usually helpful to think about long-term outcomes as the changes that can be observed at the end of, or shortly after the end of, the intervention delivery period. The pupil-level long-term outcome is usually the primary outcome captured in the evaluation and is usually related to attainment.

It is most important for the logic model to capture those short-term outcomes that are expected to equip the target population to achieve the long-term outcomes. These types of short-term outcomes are also referred to as **mediators**. For example, an intervention may be aiming to improve literacy attainment (long-term outcome) by increasing children's enjoyment of reading (short-term outcome and mediator of literacy attainment).

Programmes usually have outcomes for every type of participant they work with – for example, an intervention may have teacher- and pupil-level outcomes. When possible, it is helpful to capture in what order these outcomes are expected to occur – for instance, many programmes expect teachers to change their behaviour in the classroom, which *in turn* is expected to lead to a change in pupil behaviour. What is a short- or long-term outcome for a teacher, might be an output for the pupil. For example, if a teacher's skills to model appropriate vocabulary improve (teacher-level short-term outcome), the pupil may be exposed to a wider breadth of age-appropriate vocabulary (pupil-level output) which in turn is expected to lead to the pupil using a wider breadth of vocabulary (pupil-level short-term outcome). Arrows can be helpful to illustrate these pathways.

If required, additional levels of outcomes can be added to the logic model, such as medium-term outcomes or longitudinal (distal) outcomes.

Causal assumptions and causal mechanisms: A good ToC should also include an explanation of *why* the programme activities can be expected to lead to the anticipated change in knowledge, skills, behaviours, or attitudes. Whenever possible, these causal links should be supported by existing evidence and/or educational theory. The evaluation will further test the strength of (some of) these causal links.

In the ToC workshops, delivery teams capture causal assumptions in a 'causal assumptions log' (see Appendix 2: Assumption Logs). Delivery teams are also asked to assess the existing strength of the evidence supporting each causal assumption based on their knowledge. Evaluators may supplement information on the existing strength of the evidence on the causal assumptions via a literature review as part of preparing the protocol.

Contextual assumptions: Contextual assumptions are the underlying conditions or resources that are expected to need to exist for the programme to be implemented successfully and be effective (see Appendix 2: Assumption Logs). For example, a contextual assumption may be that teachers will be released from their usual teaching duties to attend programme training, or that schools will be able to schedule 1:1 support provided by the programme outside of usual teaching time.

Delivery teams are asked to assess i) how likely each contextual assumption is to hold in the context in which the programme will be delivered and evaluated, and ii) if the contextual assumption were not to hold (fully), how much of an impact this would have on the programme being delivered with fidelity.

Pilot evaluations that test the feasibility of a programme will often explore contextual assumptions in detail, while for efficacy or effectiveness trials, many contextual assumptions may already be established to hold. However, whenever programmes are implemented in new or more variable contexts, it is likely to be important to test whether the contextual assumptions continue to hold.

Parts of the logic model not usually covered in ToC workshops but should be discussed in IDEA workshops

Moderators: Moderators are variables that modify the form or strength of the relation between intervention and outcome. They describe for whom the intervention works, and under what circumstances.

It is often helpful to think about three types of moderators (although there is overlap across these categories):

- **Characteristics of the programme participants**, such as a teacher's prior level of qualification, or pupil gender, age, socio-economic status, disability status, having English as an additional language, or level of attainment at the start of the intervention (often captured through the baseline measure).
- **Contextual factors**, that is features of the context in which the intervention is being implemented that may influence the strength of the relation between intervention and outcome, such as the type of school, the geographic location, teachers' existing workload, support from the school's senior management team, or wider systemic or policy factors.[6]
- **Implementation factors**, particularly fidelity but also other implementation dimensions such as quality, dosage, or responsiveness. We usually expect that

[6] Contextual assumptions describe the minimum requirements expected to be in place for the intervention to be effective, while contextual moderators are elements of the context that may affect *how* effective the intervention is. There is clearly overlap between these. For example, a contextual assumption may be that pupils attend school regularly so that they are exposed to the intervention – some minimum level of attendance is likely necessary for the intervention to have any effect, and the intervention may be differentially effective for different levels of attendance.

delivering a programme with high fidelity (particularly with the expected dosage and level of quality) is more effective than delivering it with low fidelity.

Most interventions are likely to have many moderators, so we encourage specifying those that are expected to have the strongest influence on the outcome, and that will be explored within the evaluation (either only in the IPE, or both in the IPE and IE). As with other logic model components, there needs to be a clear rationale or hypothesis for why a factor is expected to influence the success of the intervention.

The EEF requires two moderators to be explored in all impact evaluations:

- **All impact evaluations are expected to include a subgroup analysis by pupils' eligibility for free school meals (FSM)** to understand whether the intervention is effective for FSM-eligible pupils. The IPE should aim to collect complementary evidence, by examining any barriers or facilitators to delivery or uptake of the intervention that may be particularly important for socio-economically disadvantaged pupils. This is included as standard in all evaluations because of the EEF's mission to help reduce the attainment gap for socio-economically disadvantaged pupils.

- **All impact evaluations are expected to explore whether implementation fidelity is associated with the primary outcome** via a complier average causal effect (CACE) analysis. This is included because it is important to attempt to understand the effect of the intervention when implemented as intended (particularly in efficacy trials). A limitation of CACE analysis is that it can only capture aspects of fidelity that can easily be measured and quantified – we call this narrower definition of fidelity 'compliance'. Complementary evidence from the IPE is therefore important to fully understand the relationship between implementation fidelity and outcomes.

It should be noted that the role of context often goes beyond a straightforward moderator relationship. For example, a programme can modify the context in which it is implemented, for example by changing behavioural norms, which in turn may affect the programme's effectiveness. Context can be considered as both dynamic and multi-dimensional (BMJ, 2021). The role of context is likely to be increasingly important in effectiveness trials and scale-up evaluations.

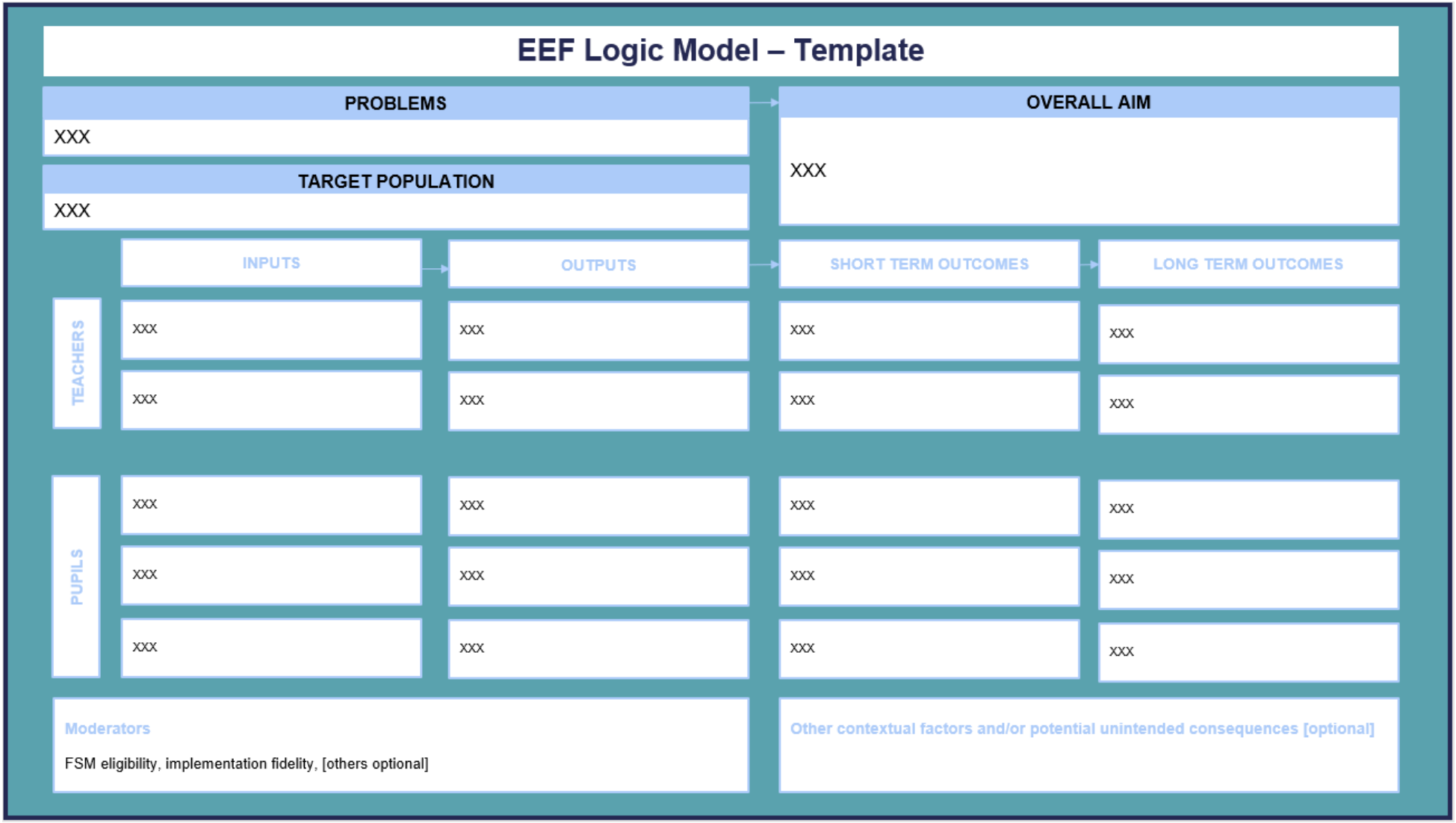
Unintended outcomes or consequences: It is important to consider early on whether the intervention might also lead to any unintended or negative consequences, so that these can be monitored or avoided[7]. For example, a programme that takes pupils out of literacy lessons to implement a new approach to studying maths may have the unintended consequence of negatively affecting literacy attainment.

Unintended consequences may occur for the programme participants, or for other teachers and pupils in the school. For example, if a programme requires teaching assistants to spend a lot of time with a small group of struggling readers, the teaching assistants may have less time to spend with other pupils in the class than they usually would. This may have the unintended consequence of negatively affecting the literacy attainment of pupils who are not participating in the intervention.

[7] Of course, unintended positive consequences may also occur.

Unintended consequences may occur for the intervention as designed or as a result of implementation challenges or adaptations made during delivery.

APPENDIX 1: EEF LOGIC MODEL TEMPLATE



APPENDIX 2: ASSUMPTION LOGS

Assumptions Log 1: Contextual Assumptions

Assumption strength - How strong do you think your assumption is?

Green – This assumption will hold in the vast majority of circumstances where the programme is delivered
Green/Amber – This assumption will hold in most of the circumstances where the programme is delivered
Red/Amber – This assumption will often not hold in the circumstances where the programme is delivered
Red – There is a good chance of this assumption not holding / do not know whether this assumption will hold or not

Assumption risk - If this assumption does not hold, how much of an impact on programme delivery with fidelity to the original design?

Green – The programme could continue to be delivered with very minor impact
Green/Amber – The programme could continue to be delivered, but the impact would be substantial
Red/Amber – The programme could continue to be delivered, but without fidelity to original design
Red – The programme could not be delivered

#	Contextual Assumption	Assumption Strength	Assumption Risk
1	XXXX	Green	Red
2	XXXX	Amber/Green	Amber/Red
3	XXXX	Amber/Red	Amber/Green
4	XXXX	Amber/Red	Green

Assumptions Log 2: Causal Assumptions

Evidence Strength - How strong is your evidence base?

Green – The evidence base is very strong. There are peer-reviewed academic studies, meta-analyses or independent experimental evaluations directly linked to the assumption.
Green/Amber – The evidence base is strong. There are academic studies or independent evaluations linked to the assumption.
Red/Amber – The evidence base is developing. There are academic studies, internal evaluations or recorded observational evidence that are adjacent to the assumption.
Red – The evidence base is limited. There is some anecdotal evidence to substantiate the assumption. You might not be aware of any evidence linked to the assumption.

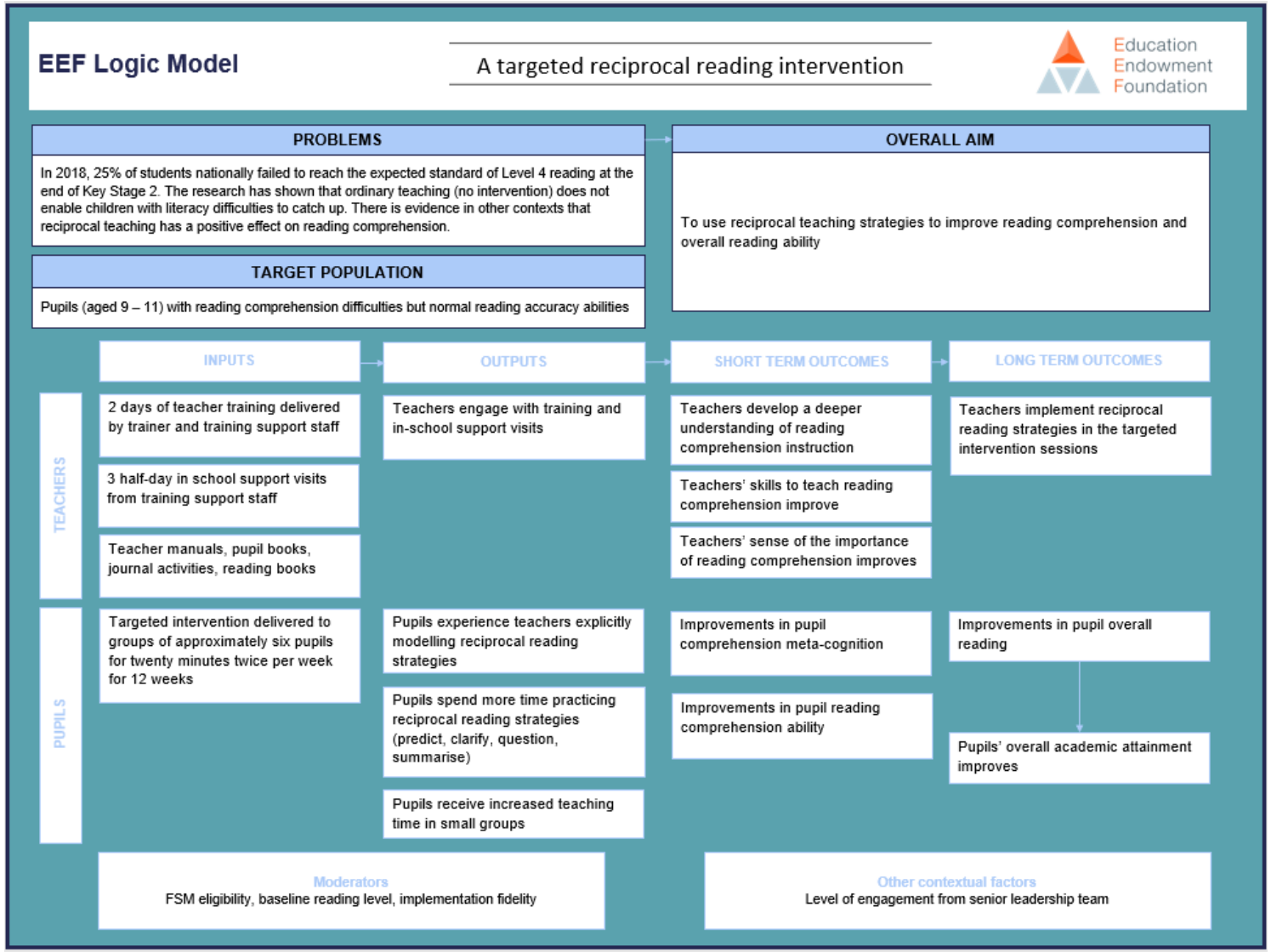
#	Assumption / Causal Mechanism	Where in the ToC does the assumption apply?		Evidence	Evidence Strength
1	XXXX	Input/Activity	Output	XXXX	Red
2	XXXX	Output	Short-term outcome	XXXX	Amber/Red
3	XXXX	Short-term outcome	Short-term outcome	XXXX	Amber/Green
4	XXXX	Short-term outcome	Long-term outcome	XXXX	Green

APPENDIX 3: TIDIER FRAMEWORK

Aspect	Description
Programme	
Why (rationale)	Rationale, theory and/or goal of essential elements of the intervention
Who (recipients)	
What (materials)	Physical or informational materials used in the intervention
What (procedures)	Procedures, activities and/or processes used in the intervention
Who (provider)	
How (format)	Format and mode of delivery
Where (location)	
When and how much (dosage)	
Tailoring (adaptation)	
How well (planned)	Strategies to maximise effective implementation

Reference: https://www.bmj.com/content/348/bmj.g1687

APPENDIX 4: EXAMPLE OF A COMPLETED LOGIC MODEL



FURTHER RESOURCES

Coldwell, M., & Maxwell, B. (2018). *Using evidence-informed logic models to bridge methods in educational evaluation*. The Review of Education, 6(3), 267–300.

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). *The program logic model as an integrative framework for a multimethod evaluation*. Evaluation and Program Planning, 24(2), 119–128.

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

Funnell, S. C., & Rogers, P. J. (2011). Purposeful program theory: Effective use of theories of change and logic models. San Francisco, CA: Wiley.

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

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.

Kaplan, S. A., & Garrett, K. E. (2005). *The use of logic models by community-based initiatives*. Evaluation and Program Planning, 28(2), 167–172.

Knowlton, L. W., & Phillips, C. C. (2012). *The logic model guidebook: Better strategies for great results* (2nd ed.). Los Angeles: Sage.

Markiewicz, A., & Patrick, I. (2016). *Developing monitoring and evaluation frameworks*. London: Sage.