FCDO Research Commissioning Centre (RCC) Evidence-Informed Policymaking (EIPM) Programme
Research Protocol

Bridging the Gap:
A cluster RCT on Strengthening Evidence Use in Ethiopian Policymaking

---

Contents

Evaluation Summary i

1. Background and Research Problem 2

2. Intervention Description 4

Rationale and Objectives 4

Recipients of the Intervention 4

Intervention Components 4

Intervention Providers 5

Timing and Duration 5

Adaptation to Context 6

3. Theory of Change/Logic of Inquiry 7

Conceptual Model and Theoretical Framework 7

Theory of Change 8

4. Impact Evaluation Design 12

a. Research Questions, Claims and Hypotheses 12

b. Methodological Approach 14

5. Process Evaluation Design 16

a. Research Questions 16

b. Methodological Approach 16

6. Cost Evaluation Design 17

7. Risks 18

8. Ethics, Safeguarding and Compliance 20

a. Ethical Considerations 20

b. Data Management and Security 20

9. Governance and Quality Assurance 21

10. Timeline 23

11. References 24

12. Annexes 26

Annex 1. Pre-Analysis Plan (Quantitative Studies) 26

1. Scope and Applicability 26

2. Study Design and Identification Strategy 26

3. Population, Sampling, and Units of Analysis 27

4. Statistical Power and Detectable Effects 28

5. Assignment Mechanism, Attrition and Missing Data 31
6. Data Collection and Measurement 33
7. Variables and Outcomes 34
8. Balance and Diagnostic Checks 36
9. Primary and Secondary Analysis Specifications 37
10. Heterogeneity, Subgroup and Compliance Analysis 38
11. Inference, Standard Errors, and Multiple Testing 38
12. Deviations and Amendments 39
13. Registration and Transparency 39
14. Qualitative sample 39
Annex 2. Data Management and Sharing Plan 40
1. Introduction and Policy Alignment 40
2. Data sources and types 40
3. Tools, software, and data formats 41
4. Data and metadata standards 42
5. Storage, backup, and access during the project 42
6. Data Security, Ethical, Legal, and Access Considerations 42
7. Data Governance and Oversight 43
8. Data Sharing and Long-Term Preservation 43
9. FAIR Compliance Summary 44
10. Resources and Responsibilities 44
Annex 3: Additional Supporting Documents 45
Annex 3a: Workplan 45
Annex 3b: Evaluation matrix 47
Annex 3c: Stakeholder Engagement and Evidence Uptake Pla 57
Core/Primary Stakeholders 61
Strategic/Secondary Stakeholders: 61
External Stakeholders: 61
Annex 3d: Added robustness checks 71

Abbreviation	Full term
AfCFTA	African Continental Free Trade Area
AEA	American Economic Association
AES / AES-256	Advanced Encryption Standard
C4ED	Centre for Evaluation and Development
CAWI	Computer-Assisted Web Interviewing
COM-B	Capability, Opportunity, Motivation – Behaviour (model)
COMESA	Common Market for Eastern and Southern Africa
CSV	Comma-Separated Values
DEFF	Design effect
DMP	Data Management Plan
DMSP	Data Management and Sharing Plan
EEA	Ethiopian Economics Association
EIPM	Evidence-Informed Policymaking
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FCDO	Foreign, Commonwealth and Development Office
FGD	Focus Group Discussion
GDPR	General Data Protection Regulation
ICC	Intracluster Correlation Coefficient
IFPRI	International Food Policy Research Institute
ITT	Intention-to-Treat
LATE	Local Average Treatment Effect
MAR	Missing at Random
MDE	Minimum Detectable Effect
NIH	National Institutes of Health
NTM	Non-Tariff Measure
OECD	Organisation for Economic Co-operation and Development
PAP	Pre-Analysis Plan
RCC	Research Commissioning Centre
RCT	Randomised Controlled Trial
RER	Rapid Evidence Review
RQ	Research Question

SD	Standard Deviation
SDC	Statistical Data Confidentiality
STATA	Statistics and Data (software)
T1	Treatment arm 1
T2	Treatment arm 2
ToC	Theory of Change
TOT	Treatment on the Treated
TREE	Transparency in Randomised Experiments and Evaluations (3ie policy)
UN	United Nations
UNEG	United Nations Evaluation Group
WTO	World Trade Organization

Evaluation Summary

Project title	Bridging the Gap: A cluster RCT on Strengthening Evidence Use in Ethiopian Policymaking
Intervention being evaluated	Evidence informed policy making in low- and middle-income countries
Principal investigator(s)/ Project lead(s)	Prof. Dr. Markus Frölich and Dr. Abule Mehare
Protocol author(s)	Abebaw Assaye, Ankhbayar Delgurchuluun, Thomas Eekhout, Markus Frölich, Johanna Gather, Atika Pasha, Nolawit Teshome
Evaluation design	Randomised Controlled Trial
Country/ies of research focus	Ethiopia
Unit of analysis	Policymakers
Anticipated sample size (at design stage)	400
Primary outcome measure and source	Increased knowledge, awareness and uptake of research (monitoring and self-reported survey data; policy documentation from ministries)
Secondary outcome measure(s) and source(s)	Increased application of evidence in policy making (policy documentation; self-reported data from survey)

1. Background and Research Problem

Ethiopia is undergoing a period of substantial reform in its trade and industrial policy framework. Trade governance combines elements of a developmental state model, such as industrial parks, export promotion, and regulatory oversight, with a shift toward more market-oriented and private-sector-led approaches. In parallel, Ethiopia is pursuing deeper regional and global integration through commitments under the African Continental Free Trade Area (AfCFTA), engagement in the Common Market for Eastern and Southern Africa (COMESA), and the resumption of negotiations for accession to the World Trade Organization (WTO), with accession targeted in the medium term. Estimates suggest that WTO membership could generate significant macroeconomic gains, but the extent to which such benefits materialise depends critically on the quality of domestic policy design, sequencing, and implementation.

Trade policy reforms in Ethiopia are inherently cross-sectoral. Agricultural commodities, including coffee, oilseeds, and horticultural products, play a significant role in export performance, linking trade policy directly to agricultural standards, competitiveness, and value chain integration. At the same time, trade-driven structural transformation interacts closely with labour and skills policies, as changes in tariffs, non-tariff measures, and market access shape firm behaviour, employment creation, and productivity. Effective coordination across trade, agriculture, labour, and industrial policy, therefore, requires policymakers to engage with complex and often technical evidence, ranging from economic analysis and firm-level data to international regulatory frameworks. Despite the availability of research relevant to these policy challenges, evidence-informed policymaking remains difficult in practice.

Challenges on evidence-informed policymaking in low- and middle-income countries are widely documented in the literature. Research consistently highlights that barriers to evidence use are not confined to individual skill gaps, but are embedded in organisational norms, institutional arrangements, and weak linkages between research producers and policy actors. Formative consultations with ministries and agencies in Ethiopia confirm the salience of these barriers: policymakers repeatedly highlighted the lack of sector-specific evidence, the difficulty of accessing relevant research, and the absence of structured spaces for dialogue with researchers. These concerns are reflected in wider literature. Building on these consultations and the literature, the challenges can be grouped into three themes: individual capability and access; institutional and systemic constraints; and the cross-sectoral patterns through which both manifest in African policymaking.

First, within the context of individual capability and access, Mnguni (2025) identifies insufficient training as a central factor limiting the use of evidence in African policymaking, while Oliver et al. (2014) emphasise the importance of the availability and accessibility of high-quality evidence. Recent work (Gerlak et al., 2023; Moallemi et al., 2023) demonstrates that collaborative and co-productive approaches increase salience and uptake. OECD (2020) and Semahegn et al. (2023) stress the need to combine capacity building with stakeholder engagement to embed evidence use in institutions. In terms of institutional and systemic constraints, Oliver et al. (2014) and Parkhurst (2017) emphasise that barriers to evidence uptake in African policymaking are often systemic rather than individual.

A third body of literature identifies cross-sectoral patterns. The same constraints recur across sectors. In agriculture, Appiah (2020) shows that while climate research in Sub-Saharan Africa generates extensive recommendations, policy uptake remains weak, with ministries struggling to integrate

findings into national agricultural strategies. Similarly, Damba et al. (2023) highlight the absence of clear frameworks for moving academic knowledge into policymaking arenas. From a labour and employment perspective, the lack of institutional channels for evidence use mirrors constraints observed in the health sector, where Nabyonga-Orem et al. (2014) argue that limited capacity and weak institutional mechanisms prevent evidence from shaping policy at scale. At the same time, Matthews (2021) (drawing on the European agricultural policy experience) illustrated that sustained uptake requires not only technical evidence but also institutionalised channels of engagement between researchers and policymakers. Yet, improving access to evidence is a necessary but not sufficient condition for its use in policymaking (Oliver et al., 2014; Parkhurst, 2017; Semahegn et al., 2023). Evidence uptake is shaped by a wider set of factors, including political incentives, organisational norms, time constraints, and the perceived relevance and credibility of research. As a result, interventions that focus solely on access are unlikely to generate sustained changes in policy behaviour without addressing these complementary constraints. This study assesses which of these factors most strongly drive the use of evidence in policy decisions in Ethiopia. It focuses on four factors that recur across the three themes above and that are tractable within a cluster-randomised design: (i) individual capability (analytical skills and confidence to appraise evidence); (ii) opportunity, defined as access to relevant evidence and structured spaces for researcher-policymaker engagement; (iii) motivation, captured through trust in research producers and perceived relevance of evidence; and (iv) organisational and institutional constraints, including time, decision-making routines, and embedded evidence culture at the directorate level. These four factors map onto the COM-B framework and the Theory of Change in Section 3, and are operationalised in the research questions in Section 4 and the evaluation matrix in Annex 3b. Political incentives and broader systemic conditions (power relations, administrative hierarchies, budget allocations) are recognised as moderating conditions rather than primary objects of measurement, and are treated as assumptions and constraints in the Theory of Change.

2. Intervention Description

Rationale and Objectives

The intervention aims to increase the uptake of research evidence in Ethiopia's trade policymaking. Low levels of evidence use are driven by a combination of individual and systemic constraints: At the individual level, policymakers often lack the technical skills and confidence to retrieve, interpret, and apply research, and may face limited incentives to do so. At the same time, ministries operate in environments where evidence use is not embedded in standard procedures, political priorities can override research inputs, and linkages between researchers and policymakers remain weak. Addressing these constraints requires more than standalone training. The intervention combines improvements in access and capacity with structured co-production and mentoring, targeting both the ability and the incentives to use evidence, as well as the relationships through which evidence becomes relevant and actionable.

Recipients of the Intervention

The intervention targets mid-level policymakers working in trade-related directorates across key institutions shaping Ethiopia's trade agenda, including the Ministry of Trade and Regional Integration, Ministry of Agriculture, Ministry of Finance, Ministry of Industry, Ministry of Labour and Skills, Ministry of Planning and Development, Ministry of Revenue, the Ethiopian Customs Commission, the Ethiopian Commodity Exchange, and the Ethiopian Investment Commission.[1] These participants are primarily analysts, technical staff, and negotiators who are directly engaged in policy processes. Senior policymakers are engaged at selected points, particularly through workshops, to ensure buy-in and support for sustained uptake within institutions. The unit of assignment is the directorate, whilst the unit of treatment and analysis is the individual officer.

Intervention Components

The intervention consists of two closely linked components. The first focuses on improving access to and capacity to use research evidence. It includes tailored training modules on econometrics, causal inference, and evidence synthesis, as well as practical guidance on accessing and interpreting academic and policy research. Participants are also provided with access to international research databases such as JSTOR, RePEc, and the World Bank eLibrary. An important element of this component is the co-development of policy-relevant research questions by participating directorates, which are then addressed through rapid evidence reviews synthesising a range of relevant evidence, including empirical studies, modelling work, and where appropriate case-based evidence on trade reforms. These reviews are conducted by the research team with structured input from a subset of participants (e.g. designated focal points). The findings from these reviews are fed back into an eight-week blended training programme combining in-person and virtual sessions, where participants engage with the evidence, discuss its implications, and identify potential applications in their own work. Training materials, evidence reviews, and outputs are stored in a centralised digital repository, and are complemented by monthly evidence digests that summarise and disseminate key findings in a format that is directly usable for decision-making.

[1] In case of changes in sample during stakeholder engagement additional units that may be considered for the design are from the Export Promotion Agency and the Trade Competition and Consumer Protection Authority.

The second component focuses on co-production and mentoring. In this component, participants engage more directly with researchers through structured collaboration. (Treated) Policymakers are organised into small research groups of approximately five to six members, each including a mentor from a research organisation and a designated "evidence champion" from within the ministry, ideally from the research team, where possible. These groups work over an extended period to produce short policy briefs and evidence digests on priority policy questions, and participate in sensemaking sessions where findings are presented, discussed, and contextualised. In addition, participants receive ongoing, on-demand mentoring to support the application of evidence in their day-to-day work, including analytical tasks, evaluations, evidence synthesis and policy drafting. Particularly, the research team will set up a response "helpdesk", to produce rapid reviews within a time frame as little as 3 days, to allow a quick response to policy makers evidence needs. This timeline may be as low as 3 hours or up to 3 weeks, depending on the complexity of the research gap being investigated (or if it is a scoping literature search and not an extensive review). This component is intended to strengthen the relevance and ownership of evidence, while embedding its use within existing workflows rather than treating it as an external input. Moreover, understanding time constraints that may arise during their workflow, this enforced partnership and linkage with researchers will ensure that two different constraints (time and organisations structures) are adapted to ease evidence availability, summarisation and uptake.

Intervention Providers

The intervention is delivered jointly by the Ethiopian Economics Association (EEA) and the Centre for Evaluation and Development (C4ED). Both organisations contribute to the delivery of both components of the intervention, combining contextual expertise with methodological and analytical capacity. Given their long-term engagement in country, EEA leads stakeholder engagement, coordination with ministries, and contextual adaptation of activities, drawing on its established networks and local presence. C4ED leads the design of training content, intervention methodological rigour, and quality assurance, particularly for the evidence synthesis. Delivery is hybrid, combining in-person workshops in Addis Ababa with virtual sessions and ongoing remote mentoring, allowing for sustained engagement across institutions.

Timing and Duration

The intervention is implemented over a 24-month period. Initial activities include participant identification and engagement, followed by the baseline data collection with all 400 study participants (between February and July 2026). Thereafter, training activities with the treatment group (approximately 220–290 participants across 36 treatment directorates, subject to confirmation at recruitment) will include a three-day econometrics and statistics training workshop, followed by the evidence synthesis, which them informs an eight-week blended learning programme (between August and November 2026). This phase (Phase I) is followed by the co-production and mentoring phase (Phase II—November 2026 to May 2027), during which participants engage in collaborative work, produce policy outputs, and receive continuous support through research mentoring and the response helpdesk. We note that within the phases I and II, certain participants remain involved at a lower intensity, for example through participation in workshops, access to evidence products, and exposure to outputs generated within the project (through conversation with junior team members or research mentor). A final phase (between May and December 2027) includes endline data

collection and a dissemination workshop bringing together policymakers, researchers, and other stakeholders.

Adaptation to Context

While the overall structure of the intervention is standardised to allow for evaluation, several elements are adapted to the Ethiopian context. Training content and evidence reviews are tailored to trade policy, with linkages to agriculture and labour, and are aligned with ongoing policy priorities such as WTO accession and regional integration. Research questions are defined by participants themselves to ensure direct relevance, and mentoring is flexible to reflect differences across ministries and directorates. At the same time, institutional elements such as the evidence repository, access to research databases, response helpdesks and the nomination of evidence champions are intended to reduce reliance on individuals and support more sustained, system-level changes in how evidence is used within policymaking.

3. Theory of Change/Logic of Inquiry

Conceptual Model and Theoretical Framework

This study examines how research evidence is used in trade policymaking and identifies practical approaches to strengthen that use. Evidence-informed policymaking (EIPM) is widely promoted as a means of improving the quality and effectiveness of public policy, yet research is not systematically integrated into decision-making processes. We conceptualise evidence use as the outcome of interactions between policymakers, researchers, and institutional environments within a policy system. To understand the behavioural processes underlying evidence use, our study draws on the COM-B model of behaviour change.

The COM-B framework posits that behaviour is influenced through the interaction of **capability**, **opportunity**, and **motivation**, which together generate, enforce and sustain behavioural outcomes.[2] Within the COM-B system, these components interact dynamically, and behaviour itself may reinforce or modify capability, opportunity, and motivation over time (Michie et al., 2011). Applied to policymaking contexts, the COM-B framework suggests that the use of research evidence depends on whether policymakers have the *capability* to interpret and critically assess research, the *opportunity* to access and apply evidence within institutional structures and processes, and the motivation to engage with research in the course of their professional work. Both individual skills and broader organisational environments shape these behavioural conditions.

Building on this behavioural perspective, the study also utilises the EIPM framework to describe the institutional and relational conditions that influence evidence use within policy systems (Sempé et al., 2025). The EIPM framework posits that research uptake depends not only on individual capabilities but also on the interactions between policymakers and researchers, the accessibility of relevant evidence, and the organisational norms governing decision-making processes. Within our study, these influences are captured through four domains that shape evidence use in policymaking. *Capabilities* refer to policymakers' skills and confidence in identifying, interpreting, and critically appraising research evidence, as well as translating findings into policy-relevant insights. *Relationships and networks* capture the degree of opportunity, trust, interaction, and mutual understanding between policymakers themselves and policymakers and researchers, which influence the perceived credibility and relevance of research. *Structures and processes* refer to institutional arrangements that affect the practical and perceived accessibility of evidence, including briefing routines, decision-making timelines, and the availability of synthesised evidence products. Finally, *evidence culture* refers to organisational norms and expectations regarding the role of evidence in routine policy work, including whether engaging with research is encouraged or rewarded within ministries and agencies. The behavioural mechanisms identified in the COM-B framework operate through these domains. *Capability* relates primarily to policymakers' analytical skills and confidence in engaging with research. Institutional structures, access to research resources, and the strength of relationships between researchers and policymakers shape *opportunity*. *Motivation* reflects both individual attitudes toward evidence

[2] Capability refers to the individual's psychological and physical capacity to perform a behaviour, including knowledge and analytical skills. Opportunity refers to the external conditions that enable or constrain behaviour, including institutional structures, social environments, and access to resources. Motivation encompasses the cognitive and emotional processes that energise and direct behaviour, including intentions, beliefs, and habitual responses.

use and broader organisational norms that signal whether engaging with research is valued within policymaking environments. Together, these elements form a conceptual model in which evidence use emerges gradually through repeated interactions between actors and institutions within the policy system. Table 1 illustrates how the COM-B and EIPM frameworks overlap and complement each other to explain research uptake.

Table 1. Integrated COM-B – EIPM frameworks

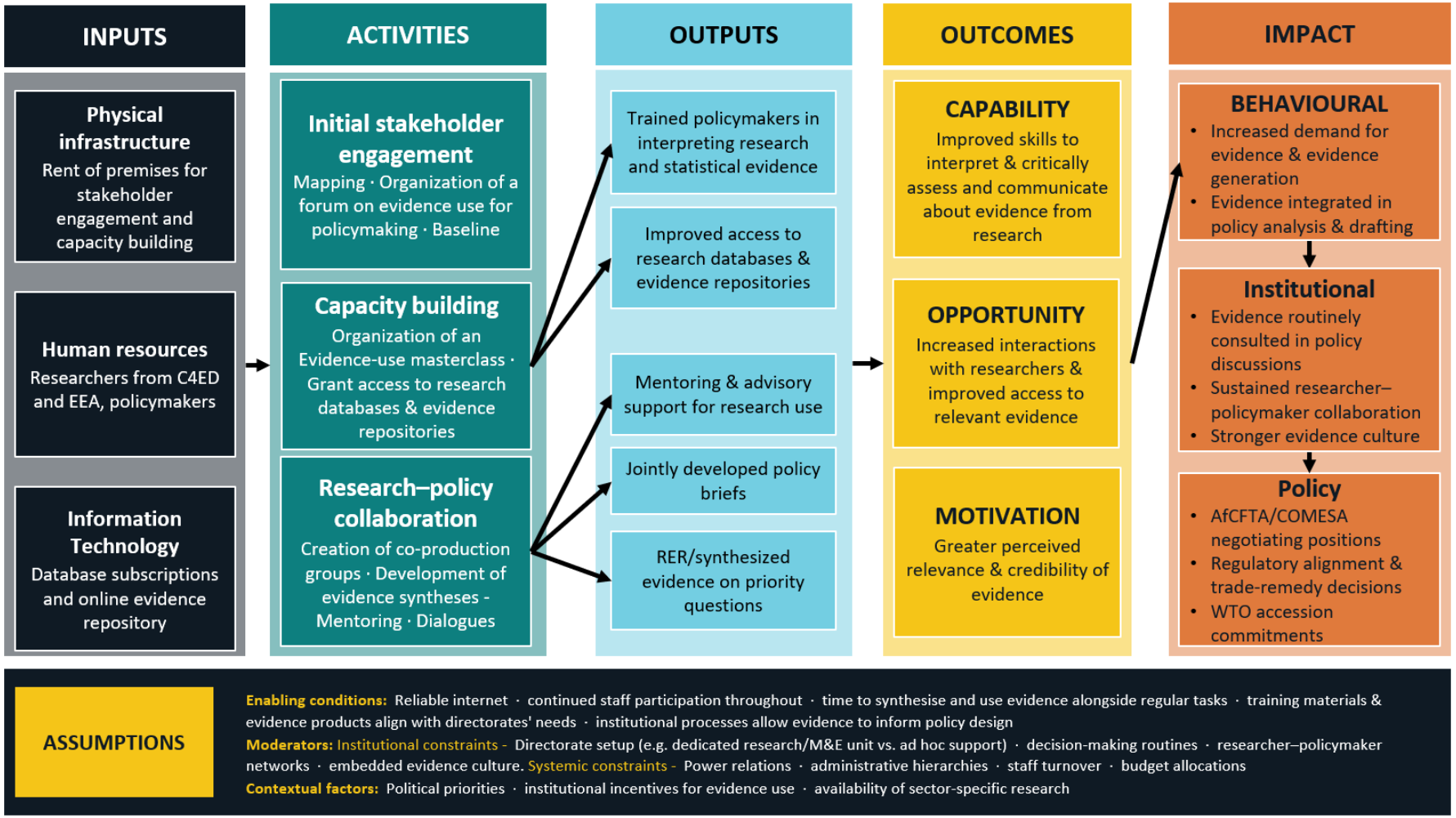
EIPM domain	COM-B component	Mechanism linking the two
Capabilities	Capability	Policymakers' skills and confidence to identify, interpret, and critically appraise research evidence and translate it into policy-relevant insights.
Relationships and networks	Opportunity (social) + Motivation	Trust and interaction between policymaker and other evidence users improve both access to evidence and the perceived credibility and relevance of findings.
Structures and processes	Opportunity (physical)	Briefing routines, decision-making timelines, and availability of synthesised evidence products determine whether research can be used for policy making.
Evidence culture	Motivation	Organisational norms and expectations signal whether engaging with evidence is valued, encouraged, and rewarded within ministries and agencies.
Behaviour: evidence use	Behaviour	The downstream outcome; sustained evidence use feeds back to reinforce capability, opportunity, and motivation across all four EIPM domains over time.

Our framework does not assume that improvements in these domains immediately translate into observable changes in policy practice. Evidence use is often constrained by time pressures, competing political priorities, and institutional incentives that influence decision-making processes. As a result, changes in analytical capability, access to evidence, or researcher–policymaker relationships may take time to manifest in observable behavioural outcomes. Evidence-informed policymaking is therefore conceptualised as a cumulative process, shaped by ongoing interactions and reinforcement within organisational settings.

Theory of Change

The intervention is designed to strengthen the use of research evidence in Ethiopia's trade policymaking by addressing the behavioural and institutional conditions described in the conceptual framework. Building on the COM-B model, the Theory of Change (ToC) (as visualised in Figure 1) assumes that evidence use is more likely when policymakers have the *capability* to interpret research, the *opportunity* to access and apply evidence within institutional processes, and the *motivation* to engage with research in the course of their work.

Figure 1: Theory of Change



Source: Own elaboration

The delivery of the intervention draws on three categories of **inputs**. The main inputs are the mobilisation of human resources. They include the researchers from C4ED and EEA who engage with policymakers, design and deliver the different activities but also the policymakers themselves, especially the evidence champions whose time engaging with training, co-production and helpdesk activities is treated as an in-kind input. Physical infrastructure inputs cover the rented premises and associated logistics used for the engagement forum and capacity-building activities. Information technology inputs cover the database subscriptions and the online evidence repository (procured and maintained by C4ED and EEA for the duration of the intervention) that condition policymakers' access to research evidence.

The intervention relies on three main **activities**. The first implies an initial engagement with the relevant stakeholders. This engagement initially implies a mapping of the latter and engaging them through an invitation to a high-level forum promoting the use of evidence in policymaking in Ethiopia (see Annex 3c for more details on the stakeholder mapping and engagement plan). The forum will introduce participants to the role of evidence in policymaking and demonstrate how it can strengthen their work. Its role is to spark enthusiasm and build momentum around evidence-informed policymaking. It will also inform the design of subsequent activities by surfacing the current state of evidence use, participants' needs, and the main barriers they face. After the organisation of the forum, the intervention will keep on engaging with the participants through two subsequent and complementary set of activities. The Capacity-building activities focus on strengthening analytical capability and improving access to research evidence through structured capacity-building activities (listed under the first component above). These activities are intended to equip policymakers with the skills and resources required to identify, interpret, and assess research findings relevant to policy questions, thereby modifying their attitude (and underlying biases) towards research and researchers. The research-policy collaboration activities focus on strengthening cooperation between policymakers and researchers. The activities intend to stimulate recurrent and structured interactions between the evidence producers/brokers (mainly C4ED and EEA) and the evidence users (directorate staff and senior decision-makers). All activities will inherently imply information flows in both directions. Policymakers will raise questions through the helpdesk and co-production sessions, researchers will respond with tailored outputs (rapid evidence reviews, policy briefs…), and users will share feedback on relevance, timing, and applicability which will help improving the research outputs for the next rounds. The co-production and mentoring components will help not only activating the information loop but also building interpersonal trust. The evidence champions will have a key role in this regard as they will represent the main intermediaries through which the loop is expected to persist.

Together, these activities are expected to generate a set of immediate **outputs**. The capacity building component will create a larger base of policymakers trained in interpreting and integrating research evidence, that have improved access to academic research, but also synthesised evidence products to deter any information overload within their time constraints. The research-policy collaboration component will lead to an increased stimulation of the "evidence" culture within directorates (through the selection of and engagement with evidence champions, as well as the joint or researcher-led creation of evidence outputs including rapid reviews, policy briefs or other policymaker requested analytical outputs within the two components).

These outputs are intended to influence the behavioural **outcomes** identified in the conceptual framework above. "As shown in the Theory of Change, each output is expected to contribute to all

three behavioural dimensions — capability, opportunity and motivation — though some pathways are more direct than others.[3] Training activities are expected to strengthen policymakers' capability by improving their skills in interpreting and critically assessing research evidence, and their ability to translate and integrate research evidence into policy, and furthermore to communicate these within relevant policy circles. Improved access to research databases, evidence syntheses, policy-relevant briefs and researchers is expected to expand policymakers' opportunity to consult research within existing policy making processes. Collaborative activities such as co-production and mentoring are expected to strengthen the underlying trust in research (and researchers), and thereby improve the perceived relevance and credibility of research evidence within policymakers. In addition, sustained collaboration between policymakers and researchers is expected to progressively improve the relevance of evidence syntheses, ensuring that the outputs produced are thematically aligned with live policy questions and delivered in time to inform decision-making.

Through the change in these outcomes, we expect impacts to materialise sequentially. First, we expect the outcomes to impact individual behaviour where policymakers will demand to integrate more evidence generation and use for policy drafting and analysis. The regular use of relevant evidence will then lead to institutional changes where directorates are routinely consulting researchers and inculcating existing evidence within their policy discussions (and policy drafts), while formalising the generation of evidence within their work cycle. This would be visible through budget allocations towards knowledge generation or translation (such as with policy briefs or synthesis of several academic studies) but also the creation of formal mechanisms for policy outreach and communication. Particularly, the creation of interactive platforms (under the helpdesk activities) for engagement between policymaker and research intermediaries (such as C4ED and EEA) would lead to evolving structures and processes that would encourage larger system level changes. At the policy level and on the longer term, these shifts are expected to translate into more evidence-informed decisions in the trade arenas currently shaping Ethiopia's integration agenda: better-analysed negotiating positions and tariff-liberalisation schedules under AfCFTA and COMESA, evidence-based choices on regulatory alignment, non-tariff measures, and trade-remedy actions, and stronger analytical underpinnings for WTO accession commitments and the sequencing of associated reforms.

Through these activities, the intervention is designed to target multiple potential **constraints** to evidence use simultaneously. The ToC operates within the institutional conditions described by the EIPM framework and distinguishes two types of constraints to evidence use: institutional and systemic constraints. *Institutional-level constraints* shape the environment in which individual capability, opportunity, and motivation translate into evidence use. These include the institutional setup of participating directorates (for example, whether a directorate has a dedicated research or M&E unit versus relying on ad hoc analytical support from line staff), prevailing decision-making routines (for example, whether policy notes are routinely cleared through an evidence-review step or drafted directly for ministerial sign-off), the strength of researcher–policymaker networks at the directorate level, and the extent to which an evidence culture is already embedded in day-to-day policy work. The intervention does not seek to reform these conditions directly, though it might lead to potential

[3] For instance, jointly developed policy briefs is directly linked to an improved access to relevant evidence (opportunity) but the collaborative process will also improve their capability to interpret and assess evidence from research (capability). In addition, the tailored nature of the policy brief will serve to increase the relevance of the evidence generated (motivation).

organisational changes by reducing the (perceived) cost (or increasing the perceived benefits) of setting up evidence-based processes, for example. These constraints will be considered in the analysis to understand their relative importance in evidence use for policy making.

Systemic constraints operate above the organisation and include power relations, administrative hierarchies, staff turnover, and budget allocations. These are likewise outside the intervention's scope and are expected to amplify or dampen observed effects: for example, high staff turnover within participating units is expected to dampen effects as trained policymakers rotate out before new routines consolidate, while stable budget allocations for analytical work that values evidence are expected to amplify effects (by giving trained staff the time, resources, and institutional backing to generate or use evidence). These systemic constraints will be treated explicitly as moderators in the analysis.

The underlying **assumptions** of the ToC include that a set of enabling conditions hold. Key enabling conditions include reliable access to Internet (i), continued participation of staff in training and engagement activities throughout the intervention (ii), time to synthesise and use evidence while still being able to perform their regular tasks (iii), training materials and synthesised evidence products align with the evidence needs of participating directorates (iv) and that the institutional processes allow policy makers to leverage the evidence found for the policy design (v). Note that we are aware that the evaluation timeline may be too short for some changes in policy practice to materialise as measurable impact. For these cases, we will report intermediate outcomes and discuss the plausibility of longer-term impact. In addition, broader contextual factors such as political priorities, institutional incentives for evidence use, staff turnover within ministries, and the availability of sector-specific research may influence how changes in capability, opportunity, and motivation translate into observable changes in policy practice.

4. Impact Evaluation Design

a. Research Questions, Claims and Hypotheses

This study is designed to generate rigorous and policy-relevant evidence on how structured interventions can strengthen the use of research in Ethiopia's trade policymaking. Anchored in the COM-B framework and the project's theory of change, the evaluation focuses on estimating causal effects of the intervention packages on policymakers' capability, behaviour, and attitudes, as well as exploring the mechanisms through which these effects occur.

The evaluation distinguishes between confirmatory (primary), secondary, and exploratory (tertiary) research questions. The confirmatory analysis focuses on estimating causal effects of the intervention across these four domains. Within this structure, behavioural uptake of evidence (RQ4) constitutes the primary outcome of interest, while capability (RQ1), opportunity (RQ2), and motivation (RQ3) represent key intermediate outcomes (mechanisms) through which behavioural change is expected to occur.

The evaluation addresses four confirmatory research questions:

- RQ1 (*Capability*): To what extent does the intervention strengthen policymakers' *capability* to interpret, critically appraise, and apply research evidence in Ethiopia's trade sector?
- RQ2 (*Opportunity*): Does the intervention expand policymakers' *opportunity* to consult research in their day-to-day policy work?

- RQ3 (*Motivation*): To what extent does the intervention strengthen policymakers' *motivation* to use evidence, operationalised as attitudes toward research, trust in research producers, and intentions to use evidence in policy processes?
- RQ4 (*Behavioural – evidence use*): Does the intervention lead to greater uptake of evidence in policy drafts, negotiations, and decision-making?

In addition, the study examines a set of secondary questions that focus on mechanisms and ToC pathways. Secondary research questions focus on unpacking the mechanisms and pathways through which these effects occur, including the role of individual skill acquisition, researcher–policymaker interaction, and organisational norms. Additional questions examine the effectiveness of evidence intermediaries, cost-effectiveness, and the conditions under which effects are sustained beyond the life of the intervention:

- Through which mechanisms, e.g., individual skill acquisition, researcher–policymaker interaction, or organisational norms, do interventions influence evidence uptake?
- How effective are trusted evidence intermediaries and embedded policy labs in bridging the gap between research producers and policymakers within Ethiopia's trade governance system?
- What are the direct and indirect costs of implementing these interventions, and how do these compare with their effectiveness and potential scalability?
- Under what conditions (institutional, political, organisational) do interventions sustain long-term evidence use, beyond the life of the study?

Tertiary (exploratory) questions examine broader institutional and system-level dynamics, including agreement on policy questions, cross-ministerial collaboration, institutionalisation of evidence-use practices, and equity and inclusion:

- *Agreement*: Does co-production foster greater consensus among policymakers on the relevance of policy questions and the types of evidence required?
- *Collaboration*: Does the intervention strengthen cross-ministerial collaboration in evidence use, particularly through joint training, shared RERs, and policy dialogues?
- *Equity and inclusion*: Do interventions affect female policymakers differently in terms of capability, opportunity, or motivation to use evidence? (important for gender balance and subgroup analysis).

The research questions are operationalised through a detailed evaluation matrix (see Annex 3b), which maps each question to specific outcomes, indicators, data sources, and methods of verification. The matrix links the conceptual framework (COM-B and EIPM domains) to measurable constructs and ensures consistency between the theory of change, data collection instruments, and empirical analysis.

The study is designed to make causal claims on the effects of the intervention packages, leveraging random assignment at the directorate level. Specifically, it identifies:

(i) the causal effect of capacity-building alone relative to business-as-usual, and
(ii) the additional causal effect of mentoring and co-production relative to training alone.

These claims focus on attribution of changes in capability, opportunity, motivation, and observed behaviour (e.g., use of evidence in policy outputs) to the intervention. The study does not aim to make causal claims about broader system-level outcomes such as final policy decisions, trade performance, or macroeconomic impacts, where attribution would not be credible.

The evaluation tests a set of core hypotheses aligned with the COM-B framework. First, in line with RQ1, capacity-building interventions are expected to improve policymakers' analytical capability. Second, consistent with RQ2, the intervention is expected to expand opportunities to access and engage with evidence within policy processes. Third, in relation to RQ3, the intervention is expected to strengthen motivation by increasing trust in research, perceived relevance, and intentions to use evidence. Finally, in line with RQ4, these changes are expected to translate into increased behavioural uptake of evidence in policymaking.

The design recognises that capability, opportunity, and motivation interact in shaping evidence use. The different components of the intervention are therefore expected to operate through distinct but complementary channels: capacity-building primarily strengthens analytical capability, while mentoring and co-production are expected to enhance opportunity and motivation, leading to differential effects across treatment arms.

b. Methodological Approach

The study adopts a mixed-methods design centred on a cluster randomised controlled trial (cRCT), complemented by the structured production of rapid evidence reviews (RERs) and an embedded process evaluation. The cRCT allows for credible causal inference on the effects of the intervention packages, while the process evaluation and RER component provide the context needed to understand how and why observed effects emerge in a complex policy environment.

The impact evaluation is the core component. Directorates within trade-related ministries and agencies serve as the unit of randomisation, while outcomes are measured at the individual policymaker level. This design reflects how policymaking is actually organised: analytical work, information sharing, and norms around evidence use are structured at the directorate level. Randomising individuals within directorates would not be credible given the high likelihood of spillovers through shared workflows, informal exchange, and joint participation in training and mentoring. A cluster design therefore improves internal validity while remaining operationally feasible.

Approximately 60 directorates are expected to participate. The expected number of directorates is based on initial mapping and discussions with partner institutions and will be finalised during recruitment. These will be allocated across three arms: around 24 directorates in the control group (business-as-usual), 16 directorates receiving the capacity-building and access package, and 20 directorates receiving the full intervention, including mentoring and co-production. Within each directorate, an estimated 6–8 policymakers will participate, yielding a total sample of roughly 350–400 individuals at baseline. The slightly uneven allocation reflects operational constraints, particularly the higher resource intensity of mentoring and co-production activities. Power calculations (documented in the PAP) indicate that this sample size is sufficient to detect policy-relevant effect sizes for primary outcomes, accounting for clustering at the directorate level. Subgroup analyses (e.g. by ministry, gender, or baseline capacity) are exploratory in nature, and the study is not specifically powered to detect heterogeneous effects across these dimensions.

To minimise differential attrition and maintain engagement, control group participants will be included in general study activities (e.g. introductory workshops and surveys) and may receive delayed access to selected materials after the evaluation period.

Randomisation is conducted at the directorate level using computer-generated random numbers and is stratified by ministry to ensure balance across institutions. Baseline data from a registration survey are used to construct cluster-level covariates (e.g., directorate size, staff composition, baseline engagement with evidence), which can further improve balance. The randomisation is coordinated with a parallel intervention implemented by IFPRI targeting senior policymakers, ensuring consistent treatment assignment for overlapping directorates and avoiding contamination across studies.

The primary outcome is policymakers' use of research evidence in their work, while secondary outcomes capture mechanisms aligned with the COM-B framework, including capability (knowledge and skills), opportunity (access to evidence and interaction with researchers), and motivation (attitudes and perceived relevance). Data are collected at baseline, midline, and endline through structured surveys, complemented where feasible by administrative and behavioural data such as participation records and policy outputs.

RERs are a central input into the intervention but are not evaluated as standalone outcomes. They function as structured, policy-relevant evidence products that reduce information constraints and anchor training and co-production activities. The process evaluation runs alongside the impact evaluation and documents implementation fidelity, participant engagement, and variation across directorates using qualitative interviews, monitoring data, and participation records. Its role is explicitly interpretive rather than causal.

This design is appropriate given the study's objectives, but it comes with non-trivial caveats. First, spillovers across directorates cannot be fully ruled out, particularly within ministries where informal exchange is common. Second, outcome measurement relies partly on self-reported behaviour, which may be subject to social desirability bias, although this is mitigated through the use of behavioural indicators where possible. Third, the intervention targets intermediate outcomes (capability, attitudes, evidence use) rather than final policy decisions, where attribution would not be credible. Fourth, heterogeneity in implementation across ministries may introduce variation that complicates interpretation, even if it reflects real-world conditions.

Alternative designs were considered but rejected. Individual-level randomisation was deemed infeasible due to high spillover risks and organisational constraints. A purely observational or quasi-experimental design would not allow credible attribution given the complexity of policymaking environments. A simpler training-only evaluation was also rejected, as it would not allow testing the added value of mentoring and co-production, which is central to the study's contribution.

A detailed Pre-Analysis Plan (PAP) specifies the empirical strategy, outcome construction, and estimation approach, and will be finalised prior to endline data collection. A Data Management Plan (DMP) outlines procedures for data collection, storage, anonymisation, and access, ensuring compliance with ethical and data protection standards. Together, these elements aim to ensure transparency, reproducibility, and robustness of the study's findings.

5. Process Evaluation Design

a. Research Questions

The process evaluation is designed to understand how the intervention is implemented in practice and through which mechanisms it influences evidence use, in line with the study's theory of change. It examines whether and how the intervention strengthens policymakers' capability, opportunity, and motivation, and how organisational and contextual factors shape these processes.

The analysis focuses on a set of core constructs derived from the theory of change. These include how evidence is used in practice and how this evolves over time; organisational norms and expectations regarding evidence use; attitudes toward the relevance and credibility of research; barriers to evidence use at individual, organisational, and systemic levels; and patterns of collaboration, including cross-ministerial interaction. In addition, the evaluation examines whether the combined intervention (training, mentoring, and co-production) generates different responses compared to capacity-building alone, and whether effects vary across participant groups (e.g. by gender).

To better understand variation in outcomes, the process evaluation includes a small number of comparative case studies contrasting high- and low-performing directorates, identified based on quantitative results. These case studies are used to examine which organisational, behavioural, and implementation-related factors help explain differences in observed outcomes across settings.

b. Methodological Approach

The process evaluation adopts a mixed-methods design combining structured monitoring data with qualitative inquiry. This approach is appropriate given that the intervention operates through multiple components and relies on behavioural and organisational change processes that cannot be fully captured through quantitative indicators alone.

Quantitative process data are drawn from implementation monitoring systems. These include attendance records for training sessions, mentoring engagements, and co-production activities; tracking of outputs such as policy briefs and evidence digests; and usage of resources such as databases or repositories where available. These data allow us to document implementation fidelity and variation in exposure ("dosage") across directorates and over time.

Qualitative data will be collected through a combination of semi-structured Key Informant Interviews (KIIs) and Focus Group Discussions (FGDs). KIIs will be conducted with selected policymakers, evidence champions, mentors, and, where relevant and applicable, external experts involved in delivery, to capture in-depth perspectives on participation, constraints, and behavioural change. FGDs will be conducted with participants at similar hierarchical levels to capture shared experiences and organisational norms while minimising power imbalances. Sampling is purposive and designed to capture variation across treatment arms, ministries, and levels of engagement.

Qualitative data collection takes place at two main points. A small number of baseline KIIs (approximately 6) will capture pre-existing practices, organisational norms, and constraints related to evidence use. Endline data collection will include around 8 KIIs and 8 FGDs to assess changes over time and document participant experiences with the intervention. In addition, the process evaluation includes a small number of comparative case studies, each combining interviews, focus groups, and document review. These case studies focus on high- and low-performing directorates (identified

based on quantitative outcomes) and examine how organisational context, engagement, and implementation fidelity shape intervention effectiveness.

Interviews and discussions focus on how participants engage with the intervention, which elements they find useful, and what enables or constrains the use of research in their day-to-day work. Particular attention is given to organisational conditions such as incentives, leadership support, time constraints, and informal norms that shape evidence use. This is complemented by a review of outputs (e.g. policy briefs and evidence digests) to assess how evidence is reflected in practice.

Qualitative data will be analysed using a thematic analysis approach. Coding combines deductive categories derived from the study's conceptual framework (COM-B and EIPM domains) with inductive identification of emerging themes. This allows for the analysis of both pre-defined mechanisms and context-specific pathways through which the intervention influences evidence use. Quantitative and qualitative data are integrated through triangulation: monitoring data identify patterns of participation and variation in implementation, while qualitative data help explain why these patterns arise and how they relate to underlying mechanisms. This is particularly important for interpreting impact evaluation results, especially where effects are heterogeneous or weaker than expected.

This approach has limitations. Qualitative data may be affected by reporting bias, particularly where participants feel pressure to present the intervention positively; this is mitigated through assurances of confidentiality and triangulation across data sources. Monitoring data may not fully capture informal interactions or spillovers within directorates. Finally, the process evaluation does not establish causality; its role is to support interpretation of impact estimates rather than to generate independent causal claims.

6. Cost Evaluation Design

A full cost evaluation or cost-effectiveness analysis is not included in the scope of this study. While the project will generate detailed evidence on the effectiveness of different intervention components, it is not designed to support a comprehensive economic evaluation. This decision reflects both conceptual and practical considerations. The primary outcomes of the study relate to intermediate behavioural changes rather than final policy or economic outcomes. Translating these intermediate effects into meaningful economic benefits would require strong assumptions and extended time horizons that go beyond the scope of the study. As a result, a formal cost–benefit or cost-effectiveness analysis would risk being speculative and potentially misleading.

Also, the intervention combines multiple components (training, evidence access, co-production, and mentoring) delivered in an integrated manner across diverse institutional contexts. Disaggregating costs in a way that allows for robust attribution to specific components would be challenging, particularly given variation in implementation intensity across directorates and over time. Finally, resource constraints and the primary focus of the study on causal impact estimation and mechanism testing limit the feasibility of collecting detailed and standardised cost data across all intervention components.

While a formal cost evaluation is outside the scope of this study, the project will document and report the costs of intervention delivery, including training, mentoring, coordination, and material inputs. This will allow for transparent reporting of resource use and provide a foundation for future cost-effectiveness or scalability analyses.

7. Risks

The study faces a number of risks related to implementation, data collection, and interpretation of results. These risks are inherent to conducting an impact evaluation within government institutions and a dynamic policy environment, and are addressed through a combination of design choices and mitigation strategies. A first key risk concerns *participation and engagement*. Policymakers operate under time constraints and shifting priorities, which may limit attendance in training, mentoring, or co-production activities. Uneven engagement across directorates could lead to variation in treatment exposure and weaken estimated effects. To mitigate this, the intervention is designed to be flexible (e.g., blended delivery formats), and engagement is reinforced through coordination with ministry focal points, scheduling aligned with policy calendars, and the nomination of evidence champions to sustain participation. Another risk relates to *attrition and incomplete data*, particularly in longitudinal survey data collection. Staff turnover, reassignment, or reduced willingness to respond over time may result in missing data and potential bias. Analytical approaches will account for attrition where necessary, including robustness checks and, where appropriate, imputation strategies.

A third concern is *spillovers and contamination across study arms*. Given that directorates operate within the same ministries, informal exchange of information and materials across units cannot be ruled out. This may attenuate estimated treatment effects. The cluster-level randomisation reduces within-unit contamination, and process data will be used to assess the extent of cross-directorate interaction. However, some degree of spillover is likely and will be acknowledged as a limitation in interpretation. A related challenge concerns *variation in implementation fidelity*. The intervention combines multiple components delivered across diverse institutional contexts, making consistent implementation difficult. Differences in leadership support, organisational culture, or administrative capacity may lead to uneven delivery and engagement across directorates. This is addressed through standardised protocols, training of implementers, and continuous monitoring, but variation is expected. The process evaluation is therefore central to documenting how implementation differs across settings and how this shapes observed outcomes.

The study also faces *measurement challenges*, particularly due to the reliance on self-reported indicators of evidence use. In a context where evidence-informed policymaking is normatively promoted, responses may be affected by social desirability bias. To mitigate this, surveys are designed to ensure confidentiality and to reduce social desirability bias through careful question framing, including the use of vignette-based assessments and indirect questioning where appropriate. Where feasible, self-reported measures are complemented with behavioural indicators such as outputs and participation records. Nonetheless, measurement error cannot be fully eliminated and will be considered when interpreting results. In addition, *political and institutional dynamics* present an important source of risk. Changes in government priorities, leadership, or ongoing reforms related to trade policy may influence both implementation and outcomes. These dynamics may also affect policymakers' incentives to engage with evidence. While such factors cannot be controlled, the intervention design allows for some flexibility in implementation, and the process evaluation will document contextual changes to support interpretation. Finally, there are risks related to *coordination with parallel interventions,* particularly the IFPRI-led study targeting senior policymakers. In addition to coordination challenges, this also introduces a potential contamination risk, as exposure to the IFPRI intervention may influence outcomes among directorates included in this study through shared individuals, hierarchical relationships, or informal information exchange. Although coordination mechanisms are in place to

align randomisation and avoid conflicting assignments, interactions between interventions may still occur and complicate attribution of effects. This risk is mitigated through joint planning and will be explicitly considered in the analysis, for example by documenting overlap and, where feasible, accounting for exposure to the parallel intervention.

8. Ethics, Safeguarding and Compliance

a. Ethical Considerations

Our study will be conducted in accordance with internationally recognised ethical standards for research and evaluation. All activities adhere to the principles set out in the UNEG Ethical Guidelines and Code of Conduct for Evaluation in the UN System, including independence, impartiality, confidentiality, avoidance of harm, and transparency.

Participation in the study is voluntary. Eligible participants will be informed about the purpose of the study, the types of data collected, and the intended use of the data before participation. Informed consent will be obtained from all participants before any data collection takes place. Participants will be informed of their right to decline participation or withdraw from the study at any stage without negative consequences.

The study follows a strict "do no harm" principle and is implemented with sensitivity to the institutional and socio-cultural context in which data collection takes place. Attention will be given to power dynamics that may arise when conducting research with government officials and to ensuring that participation in research activities is clearly distinguished from participation in intervention activities. To protect confidentiality, the collection of personal identifiers will be minimised wherever possible. Where personal data are collected, these are de-identified through the assignment of unique study identifiers prior to analysis or data sharing. No individual participant is identifiable in reports or dissemination materials unless explicit consent has been provided.

As the study involves human participants and the collection of personal data, ethical review and approval are required prior to the start of data collection. The research team is responsible for obtaining all necessary ethical approvals and authorisations from the relevant institutional review boards and national authorities before any research activities commence.

b. Data Management and Security

Data management and security procedures for this study follow the project's Data Management and Sharing Plan (DMSP), which is aligned with FAIR (Findable, Accessible, Interoperable, Reusable) principles, 3ie's TREE (Transparent, Reproducible, and Ethical Evidence) policy, Horizon Europe requirements, and EU GDPR (European Union General Data Protection Regulation) provisions. Given that the study involves human participants embedded within public institutions, particular attention is given to confidentiality, power dynamics, secure storage, and controlled data release.

To ensure the highest standards of confidentiality, the study employs a multi-layered data protection strategy appropriate for Computer-Assisted Web Interviewing (CAWI). During the active study phase, personal information collected for coordination and longitudinal tracking purposes is stored and managed within the secure SurveyMonkey platform. All participant information used for survey administration, communication, and wave tracking remains within the platform's protected infrastructure, which includes encrypted data transmission and role-based access controls. Access to identifiable participant information is restricted to designated members of the research team responsible for study coordination.

Survey response data are periodically exported from SurveyMonkey and stored on secure institutional servers at C4ED with two-factor authentication and role-based access controls. During the

analysis phase, datasets used for statistical analysis will be pseudonymized by replacing all direct identifiers with randomly generated Subject IDs. The "Master Linkage Map" connecting these IDs to participant identities will be maintained in a separate, AES-256 encrypted file stored on a secure, restricted-access server at C4ED. Access to this linkage file is strictly limited to the designated Data Manager and retained only as long as necessary to ensure longitudinal data integrity.

Prior to the public release of any data, the research team will transition the dataset from pseudonymized to fully anonymised status using Statistical Disclosure Control (SDC) procedures implemented via the sdcMicro package in R. This includes global recoding (e.g., binning of age categories and tenure), local suppression of high-risk variables, and verification of a minimum k-anonymity threshold of at least 3. These procedures reduce re-identification risk to a negligible level and ensure compliance with GDPR Recital 26.

Fully anonymised Public Use Files (PUFs), accompanying documentation, and replication code will be archived in a trusted research repository as specified in the DMSP. No direct identifiers, contact information, or linkage files will be included in any shared dataset.

9. Governance and Quality Assurance

This protocol documents the overall study design, implementation approach, data collection procedures, and ethical safeguards for the evaluation and serves as the primary reference for the conduct of the study. It is complemented by the PAP, which governs all analytical decisions, including outcome construction, estimation strategies, and hypothesis testing. While the protocol focuses on how the study is implemented, the PAP ensures transparency and pre-specification of analytical choices. Any deviations from the protocol or PAP will be documented in writing, including the nature of the change, its justification, and the timing.

The study is implemented jointly by the EEA and C4ED, with clearly defined roles to balance contextual relevance and methodological rigour. EEA is responsible for stakeholder engagement, coordination with ministries, participant recruitment, and the contextual adaptation of intervention activities. C4ED leads on research design, development of measurement instruments, data management, and statistical analysis. This division of responsibilities is intentional: it reduces conflicts of interest between implementation and evaluation functions and strengthens the credibility of the findings.

Overall governance is structured around regular coordination and oversight mechanisms. A joint project management structure ensures alignment between implementing and research teams, including scheduled coordination meetings, shared monitoring tools, and continuous communication throughout the study. Day-to-day implementation is managed by designated project leads within each organisation, while technical decisions related to the evaluation design and analysis remain under the responsibility of C4ED. External stakeholders, including participating ministries and partners, are engaged through periodic updates and consultation workshops, but do not influence analytical decisions or results.

Maintaining research independence is a core principle of the study. Although the intervention is implemented in close collaboration with government institutions, treatment assignment is determined solely through randomisation and is not subject to negotiation or adjustment by participating ministries. Data analysis is conducted independently of implementing partners and stakeholders, and findings will be reported transparently regardless of their direction or magnitude. No actor involved in

implementation or funding will have the authority to suppress, alter, or delay the publication of results. This separation is critical given the political and institutional sensitivity of policymaking environments.

Quality assurance is embedded across all stages of the study. Standardised protocols are used for training delivery, data collection, and monitoring to ensure consistency across directorates. Survey instruments are piloted prior to implementation, and data collection processes include built-in validation checks to minimise entry errors and inconsistencies. Data are regularly monitored for completeness, attrition, and irregular patterns, and any issues are addressed in real time. For key intervention components such as rapid evidence reviews, internal quality control procedures (e.g., double screening and verification) are applied to ensure methodological consistency.

For the evaluation itself, quality assurance is reinforced through pre-specification in the PAP, which reduces risks of data mining or selective reporting. Replication files and documentation will be prepared alongside the analysis to ensure transparency and reproducibility. Where feasible, code and anonymised datasets will be archived in line with the Data Management Plan.

10. Timeline

The project runs over **24 months (January 2026 – December 2027)**, structured across six phases that move from participant identification and baseline measurement through capacity building, collaborative research, and final dissemination (see Annex 3a).

Phase 0: Mapping and Inception (Months 1–6, Jan–June 2026)

The project opens with the foundational groundwork required to undertake activities over the entire two years. Following a kick-off meeting, the team will map eligible ministries and directorates, conduct stakeholder engagement and consultations, and refine the Theory of Change, to develop the study protocol and the pre-analysis plan. In parallel, ethical clearance will be prepared and submitted. Thereafter, the registration process will be undertaken to generating the sample and allow the randomisation to assign directorates into treatment and control groups to take place. Further, treated directorates will be re-randomised into two arms (T1 and T2). These participants will also be in included within the baseline workshop to take place in June. The phase culminates in a study protocol (covering the implementation plan and monitoring framework), a baseline workshop with training and data collection, cleaned baseline dataset, and registration of the pre-analysis plan.

Implementation Phase 1: Capacity Building (Months 4–11, April–November 2026)

Overlapping slightly with the inception phase, this period focuses on structured learning and the preparation for the same. Training materials will prepared and delivered during the second workshop under component one, covering basics in statistics, econometrics, and evidence synthesis. Thereafter, two RERs will be produced, which will be followed by the blended training sessions (developed by these RERs). At the end of this training, the midline quantitative and qualitative assessments will be conducted during this phase, with associated data cleaning and documentation. At the end of this phase, the evidence champions will be selected from the treated participants. This entire phase will follow the monitoring framework to allow the process evaluation to capture relevant information on implementation.

Implementation Phase 2: Collaboration and Co-Production (Months 11–21, November 2026–Nov 2027)

The second phase planned in this evaluation starts with the development of the mentor–mentee knowledge groups (anchored by evidence champions), and tailored mentoring cycles run throughout. Participants gain institutional access to academic journals and a shared evidence repository is created. Across this phase the participants will generate eight policy briefs and accompanying evidence digests, along with white papers and other research outputs where feasible. The helpdesk will be set up to ensure that policymakers can request support on evidence that will inform their policy outputs.

Phase 4 — Endline, Dissemination and Uptake (Months 17–24, May–Dec 2027)

At the end of the project, the last phase will begin with the endline quantitative and qualitative data. Thereafter, and the evidence repository will be shared with all participants, including the control group, and the data will be cleaned and documented in preparation for final analysis. The project closes with final dissemination workshops, submission of the final evaluation report and policy briefs, preparation of journal articles, and a learning event presentation for FCDO/RCC. Knowledge management and documentation activities run throughout the entire 24-month period.

11. References

Appiah, D. O. (2020). Climate policy research uptake dynamics for sustainable agricultural development in Sub-Saharan Africa. GeoJournal, 85(2), 579–591. https://doi.org/10.1007/s10708-019-09976-2

Damba, F. U., Mtshali, N. G., & Chimbari, M. J. (2023). Factors influencing the utilisation of doctoral research findings at a university in KwaZulu-Natal, South Africa: Views of academic leaders. The Open Public Health Journal, 16, e0290651. https://doi.org/10.1371/journal.pone.0290651

Gerlak, A. K., Guido, Z., Owen, G., Rodriguez McGoffin, M. S., Louder, E., Davies, J., Smith, K. J., Zimmer, A., Murveit, A. M., Meadow, A. M., Shrestha, P., & Joshi, N. (2023). Stakeholder engagement in the co-production of knowledge for environmental decision-making. World Development, 170, 106336. https://doi.org/10.1016/j.worlddev.2023.106336

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational evaluation and policy analysis*, *29*(1), 60-87.

Hjort, J., Moreira, D., Rao, G., & Santini, J. F. (2021). How research affects policy: Experimental evidence from 2,150 Brazilian municipalities. American Economic Review, 111(5), 1442–1480. https://doi.org/10.1257/aer.20190830

Matthews, A. (2021). The contribution of research to agricultural policy in Europe. Bio-Based and Applied Economics, 10(3), 185–205. https://doi.org/10.36253/bae-12322

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implementation Science, 6(1), 42. https://doi.org/10.1186/1748-5908-6-42

Mehmood, S., Naseer, S., & Chen, D. (2023). Training policymakers in econometrics. Working Paper.

Mnguni, P. (2025). The impact of training and development on employee performance in African public sector organisations: A systematic review. Journal of Management World, 2025(3), 96–111. https://doi.org/10.53935/jomw.v2024i4.1068

Moallemi, E. A., Zare, F., Hebinck, A., Szetey, K., Molina-Perez, E., Zyngier, R. L., Hadjikakou, M., Kwakkel, J., Haasnoot, M., Miller, K. K., Groves, D. G., Leith, P., & Bryan, B. A. (2023). Knowledge co-production for decision-making in human-natural systems under uncertainty. Global Environmental Change, 82, 102727. https://doi.org/10.1016/j.gloenvcha.2023.102727

Nabyonga-Orem, J., Ssengooba, F., Mijumbi, R., Kirunga Tashobya, C., Marchal, B., & Criel, B. (2014). Uptake of evidence in policy development: The case of user fees for health care in public health facilities in Uganda. BMC Health Services Research, 14, 639. https://doi.org/10.1186/s12913-014-0639-5

OECD. (2020). Building Capacity for Evidence-Informed Policy-Making: Lessons from Country Experiences. OECD Publishing. https://www.oecd-ilibrary.org/governance/building-capacity-for-evidence-informed-policy-making_ce08832d-en

Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. BMC Health Services Research, 14(1), 2. https://doi.org/10.1186/1472-6963-14-2

Parkhurst, J. (2017). The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence. Routledge.

Semahegn, A., Mengistie, B., Abdulahi, M., Gonfa, G., Haftu, A., Aschale, M., & Mihrete, M. (2023). Factors associated with health policy implementation in sub-Saharan Africa: A systematic review. Health Research Policy and Systems, 21(1), 6. https://doi.org/10.1186/s12961-023-00958-4

Sempé, L., Pande, S., Kelly, C., Humayra, W., Anwar, B., Echt, L., & Jessani, N. (2025). Evidence-informed policymaking: A conceptual framework. Narrative report articulating the conceptual framework developed for the FCDO Research Commissioning Centre programme on evidence use in policymaking. FCDO Research Commissioning Centre.

12. Annexes

Annex 1. Pre-Analysis Plan (Quantitative Studies)

1. Scope and Applicability

This Pre-Analysis Plan (PAP) governs the confirmatory quantitative analysis of the cluster randomised controlled trial (cRCT) evaluating the impact of the intervention on policymakers' use of research evidence in Ethiopia. It applies to all analyses that aim to estimate causal effects of the intervention on pre-specified primary and secondary outcomes, as well as to the comparison between treatment arms. Specifically, the PAP covers:

- estimation of intention-to-treat (ITT) effects across all treatment arms,
- estimation of treatment-on-the-treated (TOT) effects based on measured participation and compliance,
- analysis of primary and secondary outcomes derived from survey and behavioural data, and
- a defined set of pre-specified heterogeneity and subgroup analyses.

The PAP applies to outcome measurement at both midline and endline, with endline results constituting the primary endpoint for confirmatory inference. Midline analyses are treated as intermediate outcomes and interpreted with appropriate caution. The PAP does not cover exploratory analyses that fall outside the pre-specified outcomes, specifications, or subgroups. This includes, but is not limited to:

- additional subgroup analyses not explicitly defined in this document,
- alternative outcome constructions or indices developed after observing the data,
- qualitative findings from the process evaluation, and
- extended analyses of mechanisms beyond the pre-specified framework.

Such analyses may still be conducted but will be clearly labelled as exploratory and interpreted as hypothesis-generating rather than confirmatory.

Finally, this PAP focuses largely on the quantitative impact evaluation component of the study. The process evaluation, qualitative data collection, and broader triangulation exercises are governed by the study protocol and are not subject to pre-specification in the same way. Where qualitative or process data are used to interpret quantitative findings, this will be done transparently without retroactively altering the pre-specified analytical framework.

2. Study Design and Identification Strategy

The study is implemented as a cluster randomised controlled trial (cRCT), in which government directorates (clusters) are randomly assigned to one of three arms:

(i) control (business-as-usual),
(ii) capacity-building and access to evidence (T2), and
(iii) capacity-building combined with co-production and mentoring (T1).

Randomisation is conducted at the directorate level to reflect the organisational structure of policymaking and to reduce within-unit spillovers arising from shared workflows and interactions.

The identification of causal effects relies on the random assignment of clusters to treatment conditions. Under this design, and conditional on correct implementation, treatment assignment is assumed to be independent of potential outcomes. This allows differences in outcomes between treatment and control groups to be interpreted as causal effects of the intervention.

The primary estimand is the intention-to-treat (ITT) effect, which captures the impact of being assigned to a treatment arm, regardless of actual participation or compliance. This is particularly relevant in this context, where participation in training, mentoring, and co-production activities is voluntary and likely to vary across individuals and clusters. To complement ITT estimates, the analysis will also estimate treatment-on-the-treated (TOT) effects using assignment as an instrument for observed participation, recognising that these estimates rely on stronger assumptions and apply to compliers only.

Several assumptions underpin this identification strategy. First, no systematic differences between groups at baseline beyond chance imbalances, which will be assessed through balance checks. Second, limited interference between clusters (partial interference), meaning that treatment effects in one directorate do not substantially affect outcomes in another. While randomisation at the directorate level reduces the risk of spillovers, it cannot fully eliminate them given the interconnected nature of ministries. Any contamination is therefore expected to bias estimates toward zero rather than inflate effects. Third, stable treatment assignment, meaning that clusters receive the intervention as assigned, even if individual participation varies.

The study explicitly recognises that compliance with the intervention will be imperfect. Participation in training, mentoring, and co-production activities will vary across individuals and over time. As a result, ITT estimates remain the primary source of inference, while TOT estimates are interpreted as complementary and conditional on valid instrumentation.

In addition to the experimental variation, the study combines multiple sources of outcome measurement, including survey data, and behavioural indicators a. These are not used as separate identification strategies but as complementary measurement approaches. The logic is not to identify effects through multiple designs, but to strengthen inference by triangulating across different types of indicators that capture distinct dimensions of evidence use.

Finally, qualitative data from the process evaluation are used to interpret mechanisms and contextualise findings, rather than to identify causal effects. Integration across methods will therefore follow a sequential explanatory approach, where quantitative results provide the primary estimates of impact, and qualitative evidence is used to explain variation in effects, implementation fidelity, and underlying mechanisms. This integration is interpretative and does not alter the pre-specified quantitative identification strategy.

3. Population, Sampling, and Units of Analysis

The target population of this study consists of mid-level policymakers working within trade-related ministries and agencies in Ethiopia, whose roles involve policy analysis, technical advisory work, or the preparation of policy-relevant documents. These individuals are directly engaged in processes where research evidence could, in principle, inform decision-making.

The sampling frame is defined at the level of government directorates (or equivalent units) within participating institutions. These directorates represent the organisational setting in which policy

work is structured and where shared routines, workflows, and professional interactions shape how evidence is accessed and used. Eligible institutions include ministries and agencies with mandates related to trade policy and closely linked domains, including agriculture, labour, industry, and economic planning.

At the cluster level, eligibility requires that a directorate:

- has a mandate directly related to trade or trade-relevant policy processes,
- includes a minimum number of staff engaged in analytical or advisory roles, and
- provides institutional agreement to participate in the study.

Within selected directorates, the study population consists of staff members whose roles involve engagement with policy-relevant evidence. Individuals in purely administrative or support roles are excluded. Participation at the individual level is voluntary and coordinated through directorate leadership and designated focal points.

The sampling strategy is cluster-based and largely exhaustive within the set of eligible directorates. Approximately 70 directorates are expected to participate in the study. Within each directorate, 4–5 mid-level policymakers will be included, alongside 1–2 senior staff, where feasible, to support institutional engagement and uptake. This results in an anticipated baseline sample of approximately 350 individuals, subject to confirmation at the time of enrolment.

The primary unit of analysis is the individual policymaker, as outcomes such as capability, attitudes, and behaviour are measured at the individual level. However, treatment is assigned at the directorate (cluster) level, reflecting the collective nature of policymaking processes. All statistical analyses will therefore account for clustering at the directorate level.

Several differences between the sample and the broader target population are expected. First, participation is voluntary, which likely leads to positive selection of more motivated or engaged policymakers. Second, inclusion depends on institutional agreement and existing engagement with the implementing partners, which may bias the sample toward more accessible or reform-oriented directorates. Third, the requirement of a minimum number of eligible staff may exclude smaller or less resourced units. These factors limit the external validity of the findings and should be taken into account when interpreting results beyond the study sample. At the same time, the sampling approach is appropriate for the study's objective, which is to evaluate the effectiveness of structured interventions within real-world institutional settings. Internal validity is preserved through random assignment at the cluster level, even if the sample is not fully representative of all policymakers in Ethiopia.

4. Statistical Power and Detectable Effects

The study employs a cluster-based sampling strategy, with government directorates serving as the primary sampling units. All directorates that meet the predefined eligibility criteria and provide an institutional agreement to participate are considered for inclusion in the study. No sampling is conducted within ministries beyond the identification of eligible directorates; instead, the study aims to include the full set of directorates that satisfy the inclusion criteria at the time of study initiation. Within selected directorates, individual participants are drawn from the population of eligible staff members whose roles involve engagement with policy analysis, technical advisory work, or the preparation of policy-relevant documents. Individual participation is therefore nested within clusters and reflects the organisational structure of policy work. Where feasible, all eligible staff within a

directorate are invited to participate, rather than selecting a subsample, in order to capture variation in experience, responsibilities, and exposure to evidence use within the organisational unit.

Power calculations were conducted using baseline means and standard deviations reported in Mehmood et al. (2023), who studied similar interventions training policymakers in econometrics. Assuming $\alpha = 0.05$ and power of 80%, our design is adequately powered to detect policy-relevant effects. For "assessment" capability and "ratings" outcomes, the change in SD units is between 0.39 and 0.48. For policy-related behavioural outcomes, the required change in SD units wellies between 0.35 and 0.53, depending on the cluster numbers, which aligned with a medium sized effect in our sample (according to Cohen's original thresholds). With a baseline sample of approximately 350 individuals within 70 clusters (around 28 control, 19 T1, 23 T2), stratified by ministry and baseline capacity, and endline sample sizes ranging between 216 to 252 across the attrition scenarios reported in Table 2, these MDEs compare favourably with benchmarks from prior policymaker training RCTs (e.g., Perry Preschool, Abecedarian). Across the calculations, different levels of attrition have been accounted for in our power calculations (see Section 5 below for the discussion of attrition risk, mitigation, and analysis).

Table 2. Power calculations for selected indicators

Outcome	Mean	SD	ICC	Total sample endline	Total sample baseline	HH per T cluster at baseline	HH per C cluster at baseline	Target Mean	Change in SD units	% change
70 clusters; C= 28 clusters, 40% attrition, T = 42 clusters, 20% attrition										
Ratings	2.745	0.0128	0.1	252	350	5	5	2.75	0.39	0.18
Assessment	0.186	0.0126	0.1	252	350	5	5	0.19	0.40	2.69
Policy	0.262	0.0057	0.1	252	350	5	5	0.26	0.35	0.76
60 clusters; C= 24 clusters, 40% attrition, T = 36 clusters, 20% attrition										
Ratings	2.745	0.0128	0.1	216	300	5	5	2.75	0.47	0.22
Assessment	0.186	0.0126	0.1	216	300	5	5	0.19	0.48	3.23
Policy	0.262	0.0057	0.1	216	300	5	5	0.26	0.53	1.15
60 clusters; C= 24 clusters, 40% attrition, T = 36 clusters, 30% attrition (higher T attrition)										
Ratings	2.745	0.0128	0.1	216	300	5	5	2.75	0.47	0.22
Assessment	0.186	0.0126	0.1	216	300	5	5	0.19	0.48	3.23
Policy	0.262	0.0057	0.1	216	300	5	5	0.26	0.53	1.15
60 clusters; C= 24 clusters, 40% attrition, T = 36 clusters, 20% attrition (higher ICC)										
Ratings	2.745	0.0128	0.25	216	300	5	5	2.75	0.47	0.22
Assessment	0.186	0.0126	0.25	216	300	5	5	0.19	0.48	3.23
Policy	0.262	0.0057	0.25	216	300	5	5	0.26	0.53	1.15

Source: Own calculations

Three-Level Design Effect Adjustment

The power calculations presented above are based on a two-level clustering structure (individuals nested within directorates). However, given that directorates are themselves nested within ministries, the study design is more accurately characterised as a three-level hierarchy: individuals →

directorates → ministries. Failing to account for the ministry level risks underestimating standard errors and overstating the effective statistical power of the study.

To assess the level of this three-level design effect (DEFF) can affect the sample, we have included these calculations in Table 4 in Annex 3d. The DEFF captures how much the clustering structure inflates the variance of estimates relative to a simple random sample, and is calculated as follows:

$$\boldsymbol{DEFF} = 1 + (n_{ind} - 1) \times ICC_1 + n_{ind} \times (n_{dir} - 1) \times ICC_2$$

Where:

n_{ind} is the average number of individuals per directorate

n_{dir} is the average number of directorates per ministry

ICC_1 is the intra-class correlation at the directorate level, as used in the original power calculations

ICC_2 is the intra-class correlation at the ministry level (i.e. the share of total variance attributable to differences between ministries)

For our calculations, the DEFF can be interpreted as a multiplier on the variance of estimates caused by the clustered structure of the data. A DEFF of 1.0 would imply no clustering penalty, i.e. a simple random sample. A DEFF higher than this (which appears to be the case for all of our variations; see Table 4) means the variance of estimates is more that what it would be under simple random sampling, which in turn means the study has "less" effective information of an unclustered sample of the same size. In the context of this study, our DEFF values lie between 2 and 5, which has severe consequences on how well powered our study is.

This is consistent with the broader methodological literature on clustered designs. Hedges and Hedberg (2007) demonstrate that design effects in multi-level public sector settings commonly range from 1.2 to 2.5, depending on the degree of sorting across clusters and the number of units per cluster. In bureaucratic and government capacity contexts specifically, where staff within the same ministry are likely to share similar working conditions, exposure to policy mandates, and institutional culture, ICC2 values in the range of 0.05–0.15 — and corresponding DEFFs in the range of 1.3 to 2.0 — are considered plausible and consistent with published estimates. The sensitivity table in our Annex is therefore structured to span this range.

The "Effective N" in our table represents the number of statistically independent observations the sample is equivalent to, after accounting for the fact that individuals within the same directorate and ministry are more similar to one another than individuals drawn at random from the population. Given the large DEFF (between 2 and 5) our effective N falls substantially below the raw endline sample (highest is around 252), affecting the study's ability to detect the target effect sizes. Therefore, including the ministry level ICC suggests that the power of our study is weaker than the original two-level power calculations suggest.

However, within this study the expected ministry-level intra-class correlation (ICC2) is close to negligible for several reasons. First, directorates within Ethiopian ministries operate semi-autonomously, with different mandates, staffing profiles, and institutional cultures, meaning that between-directorate variance is likely to be high relative to between-ministry variance, keeping ICC2 low. Second, the outcomes being measured — ratings, assessment performance, and policy use — are behavioural and attitudinal in nature, and therefore likely to vary substantially at the individual level rather

than being heavily determined by ministry-level institutional factors. This further supports an expectation of low ICC2 relative to published benchmarks. Third, we expect limited cross-directorate spillover on the specific COM-B outcomes being measured, which reduces the risk of treatment contamination inflating within-ministry similarity artificially. Fourth, and most importantly from an estimation standpoint, the concern about ministry-level variance is largely addressed by the analytical strategy itself: as described in the estimation section, ministry fixed effects will be included in all primary specifications, directly absorbing between-ministry mean differences and the corresponding component of between-cluster variance. Under this specification, ICC2 is effectively neutralised in estimation, and the design effect collapses toward the two-level figure used in the original power calculations. As a further safeguard, the ministry-level ICC will be estimated empirically from baseline data, and if found to exceed 0.1, a three-level mixed-effects specification will be adopted as the primary inferential model and power calculations revised accordingly.

5. Assignment Mechanism, Attrition and Missing Data

Recruitment will take place at both the directorate and individual level. Eligible directorates will be identified in collaboration with participating ministries and agencies, based on their involvement in trade-related policy analysis and decision-making. Ministries will be asked to nominate directorates that meet these criteria, ensuring coverage across key institutions. Participation at the directorate level will be formalised through agreement with ministry leadership.

Within selected directorates, individual participants will be recruited through an open registration process linked to the intervention workshops. Directorate heads will be asked to nominate relevant staff, and additional eligible staff may register directly. Participation is voluntary, but directorates will be encouraged to nominate a group of mid-level policymakers (typically 6–8 individuals), alongside 1–2 senior staff to support engagement and uptake. Information collected during registration will be used both for baseline data collection and to construct cluster-level characteristics for stratification in the randomisation procedure.

Directorates constitute the unit of randomisation in this study. Eligible directorates identified across participating ministries and agencies will be randomly assigned to one of three study arms:

I. a control group (24 directorates),
II. a treatment group receiving the capacity-building and access package (T2, 16 directorates), and
III. a treatment group receiving the capacity-building package combined with co-production and mentoring (T1, 20 directorates).

Randomisation will be conducted using a computer-generated procedure implemented in Stata. Assignment will be stratified by ministry or agency to ensure balanced representation of participating institutions across treatment arms. In addition, cluster-level information collected during the participant registration process will be used to construct baseline indicators describing the composition of participating directorates, such as the number of registered participants or gender balance. Where appropriate, these indicators may be incorporated into the stratification procedure to improve balance across treatment groups.

Because a subset of directorates will also participate in a related intervention implemented by IFPRI, the randomisation procedure will be coordinated between the two research teams. For directorates that appear in both study samples, treatment assignment will be determined through a joint

randomisation process to ensure that overlapping directorates are not assigned to conflicting treatment conditions across the two studies.

Baseline information used for stratification will be collected during the registration process prior to the start of the intervention. Randomisation will be conducted after registration and baseline data collection but before any treatment-specific activities begin. For transparency and reproducibility purposes, the assignment procedure will be implemented using reproducible code with a fixed random seed. The randomisation code and assignment files will be archived as part of the study documentation.

Attrition is anticipated at two levels. Directorate-level attrition (a directorate withdrawing entirely) is expected to be limited given that participation is formalised through ministry-level agreements, but will be tracked and reported at midline and endline. Individual-level attrition (staff turnover, non-response to follow-up surveys) is the more material risk and will be monitored continuously. We will report attrition rates by treatment arm and test for differential attrition using logistic regression of an attrition indicator on treatment assignment and baseline covariates; a p-value below 0.10 will trigger the robustness analyses described below.

Differential attrition poses a risk to internal validity if attrition is systematically related to treatment status or outcomes. To mitigate this risk, control group participants will be included in key study touchpoints, such as baseline and endline workshops. Regular communication and reminders will be used to sustain participation in surveys, and contact information will be updated throughout the study period to track participants over time. In the analysis, we will assess the extent and patterns of attrition across treatment arms and test for differential attrition using baseline characteristics. Where appropriate, we will implement robustness checks, including bounding approaches and reweighting methods, to assess the sensitivity of results to attrition.

Midline and endline data collection, along with intention-to-treat analysis, will help preserve statistical power even in the presence of some loss to follow-up. Subgroup analyses by ministry, gender, and seniority will be exploratory, but the study remains sufficiently powered to detect moderate effects overall, as well as between treatment arms.

Missing data may arise due to item non-response, incomplete survey submissions, or participant attrition between study waves. The extent and patterns of missingness will first be assessed descriptively by treatment arm and key baseline characteristics. Where missing outcome data are limited and plausibly random, analyses will primarily rely on complete-case estimation. In cases where missingness is substantial or systematically related to observable characteristics, robustness checks will be conducted using multiple imputation based on baseline covariates and prior outcomes, under a Missing at Random (MAR) assumption. Attrition analyses will additionally test whether missingness differs systematically across treatment arms. Where evidence of differential attrition is detected, sensitivity analyses, including inverse probability weighting and bounding approaches, will be implemented to assess the robustness of estimated treatment effects.

Missing baseline covariates used in adjusted specifications will be imputed using pre-specified procedures to avoid unnecessary loss of observations. All imputation procedures will be conducted separately by treatment arm where appropriate, and the main findings will be compared across alternative specifications to assess robustness.

6. Data Collection and Measurement

The analysis draws on a combination of structured survey instruments, and qualitative tools. The measurement strategy is designed to capture both self-reported and observable dimensions of evidence use, recognising that no single instrument provides a complete or unbiased measure of the outcomes of interest.

The core quantitative instrument is a structured survey administered at baseline, midline, and endline using CAWI. The survey captures three domains aligned with the study's conceptual framework: capability, motivation, and opportunity. Capability is measured through vignette-based assessments and short knowledge questions requiring participants to interpret statistical outputs and assess research findings. Motivation is captured through Likert-scale items on attitudes toward evidence use, perceived usefulness of research, and willingness to engage with evidence in policy processes. Opportunity is measured through questions on access to research resources, interaction with researchers, and perceived organisational support for evidence use. The survey also includes self-reported indicators of evidence use in policy-related tasks.

The survey instrument combines elements that are adapted from existing literature on evidence-informed policymaking and policy uptake, alongside modules developed for this study to reflect the Ethiopian policy context. In particular, vignette-based assessments and questions on organisational norms and barriers are tailored to the sectors and institutions included in the study.

To complement survey-based measures, the study incorporates behavioural and administrative indicators of evidence use, where possible. These include document analysis of policy-related outputs (such as drafts, briefs, or internal notes), participation in intervention activities (e.g. training sessions, mentoring, co-production), and outputs generated within the project (e.g. policy briefs and evidence digests).[4] Where available, simple usage indicators (such as engagement with shared materials or repositories) will also be used. These data sources are used to complement survey measures and to provide additional, observable proxies for evidence use.

Qualitative instruments are used as part of the process evaluation and include semi-structured interviews, focus group discussions, and structured feedback collected during implementation. These data are used to contextualise findings, explore mechanisms and may help identify potential moderators of treatment effects at the directorate level. While they are not part of the primary quantitative estimation, they may inform the interpretation of heterogeneity in outcomes

Given that some instruments are adapted and others are newly developed, piloting will be conducted prior to full implementation. This will take place during the early phase of the project and will focus on testing comprehension, relevance, and feasibility of survey questions and tasks. Adjustments will be made where necessary to ensure that instruments are appropriate for the study context and can be implemented within the available time constraints.

[4] While our expectation is that we have access to policy-related outputs and documents at ministry level, it appears that these documents are not likely to be directorate (our unit of randomisation) specific, and may include work from control directorates and ministries. Therefore, this is one set of documents we most likely will drop from our analysis, unless we establish a treatment versus control division of these. This has also been reflected in our DSMP now. The project will rely only on outputs generated within the project for our analysis.

Data collection is aligned with the study phases and embedded in project activities. Baseline data will be collected prior to the start of the intervention, including during the initial workshop. Midline data will be collected after the initial training and evidence review phase, capturing intermediate outcomes. Endline data will be collected after completion of the intervention and will include the full survey, as well as the collection of relevant documents and qualitative data. Surveys are expected to take approximately 30–40 minutes per respondent. Administrative and monitoring data (e.g., participation records and outputs) will be collected throughout the study.

Several limitations should be noted. Self-reported measures of evidence use are subject to social desirability bias. Behavioural indicators, while more objective, are imperfect proxies and may not fully capture how evidence is used in practice. Access to internal documents may be incomplete or uneven across institutions. In addition, policymaking is a collective process, making it difficult to attribute observed outputs to individual participants. These constraints are acknowledged and addressed through the use of multiple data sources, but they cannot be fully eliminated.

7. Variables and Outcomes

This section defines the outcome variables, treatment indicators, and key covariates used in the analysis. All outcomes are pre-specified and correspond directly to the evaluation matrix (Annex 3b), which maps each research question to specific indicators, data sources, and measurement strategies. The purpose of this section is not to redefine outcomes, but to clarify how the constructs captured in the matrix are operationalised for empirical analysis.

In line with the conceptual framework and research design of this study, outcomes are organised into four primary domains of analysis, corresponding to the COM-B framework: capability, opportunity, motivation, and behavioural uptake of evidence. These domains reflect distinct but interrelated dimensions of evidence-informed policymaking. Among these, behavioural uptake of evidence constitutes the primary endpoint of interest, while capability, opportunity, and motivation represent intermediate outcomes (mechanisms) through which behavioural change is expected to occur. All four domains are analysed in the confirmatory framework, but their roles in the causal chain are explicitly distinguished.

Treatment is assigned at the directorate (cluster) level. Let $T_c \in \{0,1,2\}$denote assignment of cluster $c$to one of three study arms:

- $T_c = 0$: Control (business-as-usual)
- $T_c = 1$: Capacity-building and access package (T2)
- $T_c = 2$: Full intervention including mentoring and co-production (T1)

The primary estimands correspond to comparisons between:

- the full intervention (T1) and control,
- the partial intervention (T2) and control, and
- the full intervention (T1) and the partial intervention (T2), capturing the additional effect of mentoring and co-production.

For treatment-on-the-treated (TOT) analysis, continuous measures of treatment exposure will be constructed using monitoring data (e.g., participation in training sessions, mentoring interactions, and co-production activities), reflecting variation in engagement intensity across individuals.

Primary Outcome Domains

Capability (RQ1)

The first domain captures policymakers' capability to interpret, critically appraise, and apply research evidence. This reflects the "capability" component of the COM-B framework and corresponds primarily to the expected effects of the training component of the intervention. This domain includes indicators measuring:

- Confidence in interpreting statistical outputs, including effect sizes, coefficients, confidence intervals, and causal claims
- Ability to distinguish between different strengths of evidence (e.g. causal vs descriptive), measured through self-reports and vignette-based assessments
- Ability to translate research findings into policy-relevant implications or options

Opportunity (RQ2)

The second domain captures policymakers' opportunity to access and engage with research evidence in their day-to-day work. This includes both physical and social opportunity structures. This domain includes indicators measuring:

- Access to research databases, evidence products, and synthesis outputs
- Availability of time and budget for engaging with evidence
- Access to researchers, intermediaries, and support mechanisms
- Frequency of interaction with researchers and use of support services (e.g. helpdesks, co-production activities)

Motivation (RQ3)

The third domain captures policymakers' motivation to engage with research evidence, including both attitudes and behavioural intentions. This domain includes indicators measuring:

- Intentions to use research evidence in policymaking
- Intentions to generate or commission new evidence
- Perceptions of the usefulness, relevance, and credibility of research
- Trust in research producers and perceived responsiveness of researchers
- Perceived organisational norms supporting evidence use

Behavioural uptake (RQ4 — Primary Outcome)

The fourth domain captures behavioural uptake of research evidence in policymaking processes and constitutes the primary outcome of the study. This domain focuses on observable and self-reported use of evidence and includes:

- Use of research evidence in policy drafts, briefing notes, memos, and analytical documents
- Frequency of citing research evidence in meetings and internal discussions
- Instances of revising or refining policy proposals based on evidence
- Consultation of evidence products (e.g. RERs, databases, evidence digests)
- Frequency of engaging with researchers or commissioning evidence

Secondary Outcomes

Secondary outcomes capture intermediate changes aligned with the theory of change, particularly those related to awareness, understanding, and enabling conditions for evidence use (e.g., RQ1a and mechanism-focused research questions). Secondary outcomes are primarily measured through self-reported survey items and are analysed to provide insight into mechanisms underlying observed changes in primary outcomes. They are not the basis for confirmatory causal claims. These include:

- Awareness that evidence can shape different stages of policymaking (agenda-setting, formulation, revision)
- Understanding of what constitutes meaningful evidence-informed policymaking beyond symbolic use
- Agreement that ministries should commission or co-produce evidence when gaps exist
- Perceived interaction with researchers and intermediaries
- Perceived organisational support, including leadership, peer norms, and routines

Exploratory Outcomes

A set of additional outcomes capture broader institutional and system-level effects of the intervention. These outcomes are analysed descriptively and interpreted as hypothesis-generating rather than confirmatory. These correspond to exploratory research questions and include:

- Cross-ministerial collaboration in evidence use
- Institutionalisation of evidence-use practices (e.g., evidence champions, repositories, review routines)
- Consensus among policymakers on policy priorities and evidence needs
- Sustainability of evidence use under institutional or political constraints
- Gender differences in capability, access to networks, and evidence use behaviour

Each indicator specified in the evaluation matrix is treated as a distinct outcome variable in the main analysis. Where multiple indicators capture related constructs within a domain, outcomes may be grouped into families for the purpose of multiple testing adjustments. Any aggregation of indicators into composite indices will be conducted only as a robustness check and will be clearly distinguished from the primary analysis. The main specifications will include a set of pre-treatment covariates to improve precision. Covariates are included to improve statistical efficiency and do not affect identification, which relies on random assignment. These include:

- Baseline values of outcome variables (where available)
- Individual characteristics such as gender, role, and years of experience
- Cluster-level characteristics, including ministry affiliation and directorate size

8. Balance and Diagnostic Checks

Baseline balance between treatment groups will be assessed to verify the integrity of the randomisation. Balance will be examined across a set of pre-treatment characteristics at both the individual and cluster level. These include baseline measures of key outcomes (where available), individual characteristics (e.g., role, experience, gender), and cluster-level variables (e.g., ministry or agency affiliation, cluster size). Comparisons will be conducted across all treatment arms. Balance assessment will rely on mean comparisons across groups, standardised differences and joint significance tests of baseline

covariates. Any observed imbalances will be documented but not interpreted as evidence of flawed randomisation. Where imbalances are substantial, the corresponding variables may be included as controls in regression specifications to improve precision, rather than to correct bias.

In addition to baseline balance, the analysis will include diagnostic checks related to attrition and missing data. This includes comparing baseline characteristics of attritors and non-attritors, and respondents with complete and incomplete data. These comparisons will be conducted across treatment arms to assess whether attrition or missingness is systematically related to treatment status. Overall, these diagnostic checks are intended to provide transparency on the realised sample and to support interpretation of results, rather than to inform model selection or post hoc adjustments.

9. Primary and Secondary Analysis Specifications

The impact evaluation will adopt an intention-to-treat (ITT) framework, comparing outcomes between treatment and control clusters as randomised. Standard errors will be adjusted for clustering at the directorate level, with stratification variables (e.g., ministry) accounted for in estimation. The primary model will take the form of:

$$Y_{it} = \alpha + \beta_t T_c + \gamma X_{i0} + Y_{i0} + \epsilon$$

- where Y_i denotes outcomes for individual i at midline (t=1) or endline (t=2).,
- β provides the ITT estimate.
- T_c is the treatment assignment (at the cluster level),
- X_{i0} are baseline covariates, and
- Y_{i0} is the baseline value of the outcome at baseline (if available)
- ϵ is a random error term.

Given that the design nests individuals within directorates within approximately 10 participating ministries, we account for ministry-level variation in two ways. First, randomisation is stratified by ministry, and ministry fixed effects will be included in all primary specifications, absorbing between-ministry mean differences and the corresponding component of between-cluster variance. Second, as a robustness check, the primary model will be re-estimated as a three-level mixed-effects specification with random intercepts at both the directorate and ministry levels, and the ministry-level ICC will be reported empirically using baseline data. If the empirical ministry-level ICC is non-negligible (a threshold of >0.1 will be used), the three-level specification will be treated as the primary inferential model and standard errors and power calculations will be revised accordingly.

In this exercise, an important factor to consider is the degree of compliance of the treatment observations. As it is possible that not all observations will fully comply with the treatment assignment, we will also estimate the local average treatment effect (LATE) to measure the effects on the compliers using a Two-Stage Least Squares (2SLS) procedure. The procedure follows a two-step approach, where one predicts the compliance with the random selection into the treatment in the first stage:

$$C_i = \alpha + \beta_t T_c + \gamma X_{i0} + \epsilon$$

where C_i captures the extent of compliance with the assigned treatment. Given that participation in the intervention may vary in intensity (e.g., attendance at training sessions, engagement in mentoring or co-production activities), compliance will be measured using indicators of treatment exposure

rather than a purely binary measure. Then, the estimated compliance $\hat{C}_i$ is used to estimate the impact on the outcome of interest in the second stage equation:

$$y_{it} = \beta_t \hat{C}_i + \gamma X_{i0} + Y_{i0} + \epsilon$$

The estimations will be conducted in Stata. Subgroup analyses will be structured by ministry type, baseline evidence-use capacity, and participant gender, to explore heterogeneity of effects, where possible. Findings from the impact evaluation will be triangulated with complementary components of the study to strengthen internal and external validity. The institutional mapping exercise will help interpret heterogeneity across ministries by identifying governance structures and decision-making pathways that may condition treatment effects. Finally, the process evaluation aims to explain variations in outcomes by documenting fidelity and identifying barriers. Triangulating these components will allow us to generate nuanced insights into *what worked, why, and for whom*. The details on the process evaluation are in the next section.

10. Heterogeneity, Subgroup and Compliance Analysis

The primary analysis focuses on average treatment effects. In addition, a limited set of heterogeneity analyses is pre-specified to explore whether effects vary across key dimensions motivated by the ToC. Pre-specified heterogeneity will be examined along the following dimensions:

- **Baseline capability** (e.g., knowledge or assessment scores), capturing whether effects differ by initial ability to engage with evidence,
- **Gender**, to assess whether impacts differ across male and female policymakers, and
- **Institutional context**, proxied by ministry or agency, reflecting variation in organisational environments.

Heterogeneity will be analysed by interacting treatment indicators with the relevant baseline characteristics within the main regression framework. These analyses are intended to test whether the intervention is more effective for certain groups, but are expected to be lower powered than the main estimates and will be interpreted cautiously.

Compliance with the intervention is expected to vary across individuals and clusters, particularly for components requiring sustained engagement, such as mentoring and co-production. Participation will be measured using monitoring data (e.g., attendance, engagement in activities, outputs produced). To assess the effect of actual participation, treatment-on-the-treated (TOT) estimates will be obtained using assignment to treatment as an instrument for participation. These estimates identify the effect for compliers under standard instrumental variable assumptions and are interpreted as complementary to the primary ITT results.

11. Inference, Standard Errors, and Multiple Testing

All statistical inference will account for the clustered nature of the design. Standard errors will be clustered at the directorate level, corresponding to the unit of randomisation. Where stratification is used in the assignment process, stratification variables will be included as controls in regression specifications. Considering the presence of multiple outcomes, the analysis will use a combination of approaches to address multiple hypothesis testing. First, outcomes are grouped into conceptually related domains (e.g., evidence use, capability, motivation, opportunity), and where appropriate, combined into composite indices to reduce dimensionality. Second, adjustments to p-values may be

applied within outcome families using standard procedures (e.g., Holm or false discovery rate corrections), particularly for secondary outcomes and heterogeneity analyses.

The primary outcome (evidence use) is defined ex ante and will be the main focus of inference. Secondary outcomes and subgroup analyses will be interpreted with greater caution, particularly where multiple comparisons are involved.

Uncertainty will be assessed using confidence intervals and p-values, with emphasis placed on effect sizes and consistency across related measures rather than statistical significance alone. Given the limited number of clusters and potential imprecision, null results will be interpreted carefully and not taken as definitive evidence of no effect.

Composite outcomes and indices will be constructed using pre-specified rules. Where indices are used, individual components will be standardised and combined such that higher values consistently reflect outcomes in the intended direction. Sensitivity checks using alternative constructions or individual components will be reported where relevant.

12. Deviations and Amendments

Any deviations from this Pre-Analysis Plan will be documented transparently and justified in all study outputs. This includes changes to outcome definitions, estimation specifications, sample restrictions, or analytical procedures that occur after this plan has been finalised. Deviations will be recorded with a clear description of the change, the rationale for the deviation, and the timing of the decision relative to data access and analysis. All analyses will be clearly labelled as either pre-specified or non-pre-specified (exploratory) in reporting. Where deviations affect core outcomes or specifications, results based on both the original and amended approaches will be reported where feasible, to allow comparison. Minor adjustments that do not affect the interpretation of results (e.g., coding corrections, data cleaning decisions consistent with this PAP) will not be treated as substantive deviations but will be documented in replication materials.

13. Registration and Transparency

This PAP will be registered prior to randomisation and baseline data collection, alongside the study registration in the AEA RCT Registry, consistent with 3ie's TREE policy. The registration process has been initiated and, with ethical clearance now secured, registration will be completed before any outcome data analysis begins (conditional on approval from donor). The PAP registration will occur after finalisation of the study design and instruments. All analysis code and, where possible, de-identified data will be made available in a public repository following completion of the study, in line with the Data Management Plan and applicable data protection requirements. Study outputs will clearly distinguish between pre-specified and exploratory analyses. Any updates to the PAP after initial registration will be versioned and archived to ensure transparency over time.

14. Qualitative sample

As a standard practice in qualitative research, qualitative sample size will be determined by data saturation. However, based on C4ED's experience, a total of 30 qualitative interviews (18 KIIs and 12 FGDs) is proposed. With each FGD including an average number of 8 participants, the total number of qualitative interview participants will be 114.

Table 3: Qualitative sample

Phase	Method	Respondents	Number of sessions	Participant per session	Total number of participants
Baseline with potential treatment group					
Baseline	KIIs	Purposively selected individuals from the treatment group	6	1	6
Endline with treatment group (T2)					
Endline	KIIs	Purposively selected individuals across diverse institutions and directorates	8	1	8
Endline	FGDs		8	8	64
Endline case studies with most and least successful cases					
Case study 1	KIIs	Institution/directorate with most successful intervention outcome	2	1	2
	FGDs		2	8	16
Case study 2	KIIs	Institution/directorate with least successful intervention outcome	2	1	2
	FGDs		2	8	16
KIIs			18	-	18
FGDs			12	8	96
Total			**30**	**-**	**114**

Source: Own elaboration

Annex 2. Data Management and Sharing Plan

1. Introduction and Policy Alignment

This Data Management and Sharing Plan (DMSP) describes how data generated by this study will be collected, managed, preserved, and shared. The plan follows FAIR (Findable, Accessible, Interoperable, Reusable) principles and 3ie's TREE (Transparent, Reproducible, and Ethical Evidence) policy, is informed by the National Institute of Health (NIH) Data Management and Sharing Policy, and Horizon Europe data management requirements, and is aligned with European Union (EU) General Data Protection Regulation (GDPR) provisions.

The study is a clustered randomised controlled trial involving human participants embedded within public institutions in Ethiopia. Accordingly, the plan balances open science objectives with ethical obligations, data protection requirements, and institutional constraints. The overarching aim is to maximise the transparency, reusability, and learning value of the data, guided by the principle "as open as possible, as closed as necessary," while safeguarding confidentiality, informed consent, and legitimate institutional interests.

2. Data sources and types

The project will not use any existing data as none exist currently. All the data used for the project and shared with the public are expected to be generated as new data.

2.1 Data type

The project will provide human-subjects research data related to evidence-informed policymaking in Ethiopia's trade (and related to trade) governance system(s). These data are generated at the individual and aggregated directorate (cluster) levels and include the following types:

- Quantitative survey data collected from approximately 400 policymakers across 60 directorates in trade(-related) ministries and agencies. These data include participants' demographic and professional background characteristics, as well as measures of analytical capability, knowledge, motivational orientation, behavioural intentions, and reported engagement with research evidence in trade policy processes.

- Vignette-based assessment data testing participants' ability to interpret and apply research findings in policy-relevant scenarios.
- Monitoring data, including participation records for trainings, workshops, mentoring activities, response helpdesk and rapid evidence reviews (RERs).
- Qualitative data from individual interviews and group discussions with policymakers, supporting process evaluation and interpretation of quantitative findings.

2.2 Data to be preserved and shared

In line with the above-mentioned policies and guidelines, the following will be preserved and shared:

- Fully de-identified raw and clean quantitative datasets (provided in both .dta and open .csv formats)
- Aggregated and directorate-level indicators and indices
- Survey instruments, vignettes, protocols, and data dictionaries
- Annotated statistical replication code (provided in STATA do files and/or R scripts)
- Documentation (in section 8) describing data collection, processing, and anonymisation
- The qualitative sampling frame will be documented in a dedicated sampling log that records the participant selection criteria and sampling strategy (i.e., purposive and using maximum variation). It will include the relevant characteristics used for selection without retaining the participants' personally identifiable information
- MAXQDA project files (.mx format) will be archived at the close of the qualitative data collection and analysis phase. A clean, finalized version of the project file, including all applied codes, memos, and coded segments, will be stored in the secure project repository.
- A codebook export will be provided, to ensure accessibility without the need for of software licensing. Aggregated qualitative findings will be presented in structured thematic summary, organized by research questions and emerging themes. Each section of the summary will include a description of the theme, illustrative (and anonymized) excerpts, and an indication of the prevalence of the theme across respondents. Summaries will be produced in Word format and included as annexes or integrated sections in the final report.

The following data will not be shared openly:

- All direct or indirect identifiers (e.g., names and email addresses) of the participants
- Raw qualitative transcripts containing potentially identifiable information
- Confidential or draft-only government documents subject to institutional or legal restrictions

3. Tools, software, and data formats

Data will be managed and shared using widely accessible and commonly used tools and formats, including:

- Web-based survey platform (SurveyMonkey) for conducting quantitative online surveys
- STATA, and R/RStudio for quantitative analysis
- CSV formats for tabular datasets
- MAXQDA for coding and analysing the qualitative interviews
- TXT/PDF formats for documentation

All code will be documented to support reproducibility.

4. Data and metadata standards

The project follows recognised social science and policy research standards, consistent with FAIR principles and DataCite Metadata schema. The metadata includes:

- Harmonised variable naming conventions, coding schemes, and codebooks
- Unique, anonymised identifiers for individuals and directorates
- README files describing data structure, processing steps, anonymisation procedures, and limitations

These standards ensure consistency, interpretability, and interoperability across datasets and users.

5. Storage, backup, and access during the project

During the project, quantitative survey data and associated metadata will be stored within the SurveyMonkey platform, which serves as the primary system for administering the online surveys. All other project data, including qualitative materials, analytical datasets, and documentation, will be stored on C4ED's institutional data infrastructure with two-factor authentication and role-based access controls. C4ED will maintain responsibility for the primary storage and management of project data, with access limited to authorised project researchers.

6. Data Security, Ethical, Legal, and Access Considerations

6.1 Informed Consent and Confidentiality

All participants will be required to provide informed consent covering:

- Their participation in research activities
- Use of their de-identified data for analysis and dissemination
- Sharing of their data for future research purposes, subject to safeguards

To ensure the highest standards of confidentiality, the study employs a multi-layered data protection strategy appropriate for substantial sensitive data, which includes direct and indirect identifiers. During the active study phase, personal information collected for coordination and longitudinal tracking purposes is stored and managed within the secure SurveyMonkey platform. All participant information used for survey administration, communication, and wave tracking remains within the platform's protected infrastructure, which includes encrypted data transmission and role-based access controls. Access to identifiable participant information is restricted to designated members of the research team responsible for study coordination.

During the analysis phase, datasets used for statistical analysis will be pseudonymized by replacing all direct identifiers with randomly generated Subject IDs. The "Master Linkage Map" in .txt format that connects these IDs to participant identities will be maintained in a separate, AES-256 encrypted file, stored on a secure, restricted-access server at the C4ED. Access to this decryption key is strictly limited to the designated Data Manager.

Prior to the public release of any data, the research team will transition the data from pseudonymized to fully anonymized status using Statistical Disclosure Control (SDC) via the sdcMicro package in R. This process involves the rigorous application of global recoding—such as the binning of ages and geographic locations—and local suppression to achieve a k-anonymity threshold of at least 3. This

ensures that the risk of re-identification is minimised to a negligible level, satisfying the requirements for anonymization under GDPR Recital 26.

6.2 Access and Reuse Conditions

- De-identified Public Use Files (PUFs) derived from the quantitative survey and vignette data will be made openly available through the designated repository.
- Data subject to confidentiality, legal, or institutional constraints will be shared in aggregated or restricted form where necessary.
- Any access restrictions will be clearly documented in repository metadata.

7. Data Governance and Oversight

Oversight of data management and sharing will be jointly managed by the implementing partners:

- A designated Data Manager/Research Lead at the C4ED will oversee data quality, anonymisation, documentation, and repository submission.
- The Ethiopian Economics Association (EEA) will ensure compliance with local ethical approvals and institutional data-sharing agreements.

Compliance with this DMSP will be reviewed semi-annually during implementation and prior to public data release.

8. Data Sharing and Long-Term Preservation

8.1 Repositories

Research data and associated metadata will be archived in Zenodo, an EU-supported, trusted research data repository operated by CERN and compliant with Horizon Europe open science requirements. Zenodo ensures long-term preservation, assigns persistent identifiers, and supports rich metadata, making it suitable for sharing social science and policy-relevant research data.

Analysis scripts, replication code, and methodological documentation will be maintained in a public GitHub repository to support transparency, collaborative development, and version control.

8.2 Findability and Identifiers

All datasets and archived code releases will be assigned persistent digital object identifiers (DOIs) through Zenodo. Each dataset will be accompanied by standardised metadata and descriptive keywords to support discovery through Zenodo's search functionality and external indexing services.

The bidirectional linking between Zenodo records and GitHub repositories will ensure clear traceability between data, code, and analytical outputs, supporting reproducibility, citation, and long-term reuse.

8.3 Timelines

- De-identified quantitative datasets, documentation, and replication code will be made publicly available no later than the time of publication.
- Aggregated qualitative outputs will be shared on the same timeline.
- Data will be preserved for a minimum of 5 years, or longer where repository policies allow.

9. FAIR Compliance Summary

9.1 Making Data Findable

All datasets will be assigned DOIs. Rich metadata will be provided, including study descriptions, geographic and institutional coverage, units of analysis, and variable definitions in machine-readable formats (.txt) to enable harvesting and indexing.

9.2 Making Data Accessible

Data will be deposited in a trusted repository, Zenodo, that ensures long-term preservation and identifier resolution. De-identified datasets will be openly accessible through free and standardised access protocols built in Zenodo. Metadata will be openly available under CC0 and will remain accessible even if specific datasets are restricted or withdrawn.

9.3 Making Data Interoperable

Data will be shared in non-proprietary, widely used formats. Standard social science vocabularies and classifications will be applied. Project-specific indicators will be clearly defined and mapped to commonly used concepts. Datasets will include references to related project outputs and relevant external data sources where applicable.

9.4 Increasing Data Reuse

Comprehensive documentation will be provided, including README files, codebooks, and annotated analysis code. Data provenance, quality assurance processes, and version control will be documented to support validation and reuse by third parties beyond the life of the project.

The qualitative codebook will be documented for reuse during endline data collection. However, the codebook used during endline might slightly differ as a result of the development of new themes after baseline, in which case the changes will also be clearly indicated.

10. Resources and Responsibilities

No additional resources beyond the existing institutional data management infrastructure at C4ED and standard research staff allocations are required to implement the data management procedures and ensure compliance with FAIR principles.

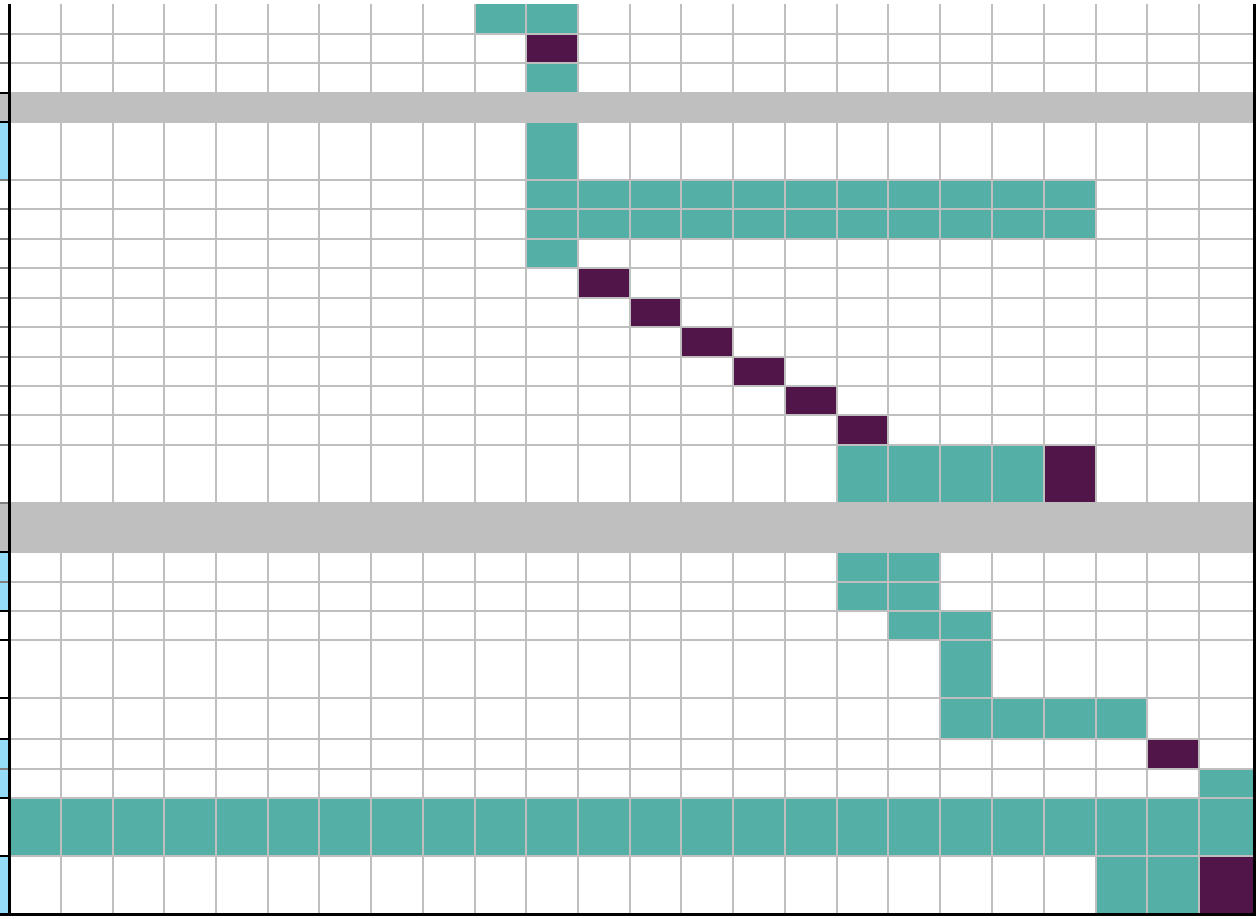
Annex 3: Additional Supporting Documents

Attach any other supporting documents as relevant, including data collection instruments, consent forms, risk registers, workplans, etc.

Annex 3a: Workplan

The current timeline for the project is as follows:

Project Activities	Project duration																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
	Jan-26	Feb-26	Mar-26	Apr-26	May-26	Jun-26	Jul-26	Aug-26	Sep-26	Oct-26	Nov-26	Dec-26	Jan-27	Feb-27	Mar-27	Apr-27	May-27	Jun-27	Jul-27	Aug-27	Sep-27	Oct-27	Nov-27	Dec-27
Project Activities																								
Mapping and Inception phase (Phase 0)																								
Kick-off meeting																								
Mapping ministries and directorates within																								
Stakeholder engagement and consultation																								
Refinement of Theory of Change																								
Ethical clearance preparation and submission																								
Registration and randomisation of directorates into T and C groups																								
Research protocol (including - implementation plan and monitoring framework)																								
Preparation of baseline workshop (and survey)																								
Baseline workshop and tailored training sessions (including baseline and training data collection)																								
Dataset cleaning and prep																								
Development and registration of pre-analysis plan (not incl. baseline data) Update in August																								
Implementation Phase 1																								
Preparation of training materials																								
Training on statistics and introduction to research																								
Rapid evidence review 1 and 2																								
Blended training																								
Midline qualitative assessment																								
Midline quantitative survey																								

Midline data cleaning and documentation	
Cleaned midline dataset	
Selection of evidence champions	
Implementation Phase 2	
Formation of mentor-mentee knowledge groups (including evidence champions)	
Tailored mentoring	
Institutional access to journals	
Creation of evidence repository	
Policy brief 1 & evidence digests	
Policy brief 2 & evidence digests	
Policy briefs 3 and 4 & evidence digests	
Policy brief 5 & evidence digests	
Policy briefs 6 and 7 & evidence digests	
Policy brief 8 & evidence digests	
Development of white papers and other research output (where possible and on demand)	
Endline and Dissemination phase	
Endline quantitative survey	
Endline qualitative assessment	
Endline data cleaning and documentation	
Sharing evidence repository with all participants (control group)	
Analysis and preparation of final report	
Final report and policy brief	
Dissemination event (workshops)	
Knowledge management & documentation (throughout)	
Finalisation of research papers for research publications	

Source: Own elaboration

Annex 3b: Evaluation matrix

The following evaluation matrix is based on the project ToC.

	#	Questions	Hypothesis	Pathway of change (RCC framework)	Topic (COM-B)	Sub-topic	ToC Level	Indicator	Means of verification	Source	Analysis	Comparison
Primary	RQ1	To what extent does the intervention strengthen policymakers' capability to interpret, critically appraise, and apply research evidence in Ethiopia's trade sector? (Capability)	Structured training, co-production and mentoring improves policymakers' **capability** (to i**nterpret, critically appraise and apply** research evidence)	**Capabilities**	**Capability**	**Interpretation**	Output	# of capacity building sessions held	- Midline	Monitoring data	Descriptive	
					Capability	**Interpretation**	Output	Participation to capacity building sessions	- Midline	Monitoring data	Descriptive	
					Capability	**Interpretation**	Outcome	Confidence in interpreting a coefficient from a test or regression accurately	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
					Capability	**Interpretation**	Outcome	Can interpret a coefficient from a regression accurately	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
					Capability	**Appraisal**	Outcome	Ability to distinguish strength of evidence: e.g. strong (causal) evidence from descriptive evidence	- Baseline - Midline - Endline	Self-reported or Experimental (Vignette) quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
					Capability	**Application**	Outcome	Ability to translate research findings into policy-relevant	- Baseline -	Self-reported quantitative and	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C;

							implications or options.	Mid-line - End-line	monitor-ing; Quali-tative		T2 vs T1
RQ2	Does the intervention expand policymakers' opportunity to consult research in their day-to-day policy work? (Opportunity)	The intervention improves research access and attitudes of policy-makers (definition of opportunity) Granting access to research databases, evidence syntheses, policy-relevant briefs and mentoring through researchers expands policymakers' opportunity to collaborate with other evidence users/brokers/intermediaries Granting access to research databases, evidence syntheses, policy-relevant briefs and mentoring through researchers expands policymakers' opportunity to use and consult research within existing policy making processes	**Structures & Processes**	**Opportunity (physical)**	**Use/Generation**	Output	# of synthesised-evidence products developed for the directorate	- Mid-line - End-line	Monito-ring data	Descrip-tive	
			Structures & Processes	**Opportunity (physical)**	**Access**	Output	Has an account to access research databases	- Base-line - Mid-line - End-line	Self-re-ported quantita-tive	Infer-ence	T vs C ; T1 vs C; T2 vs C; T2 vs T1
			Structures & Processes	**Opportunity (physical)**	**Access**	Out-come	Awareness of existence evidence repositories and archives	- Base-line - Mid-line - End-line	Self-re-ported quantita-tive	Infer-ence	T vs C ; T1 vs C; T2 vs C; T2 vs T1
			Structures & Processes	**Opportunity (physical)**	**Attitudes**	Out-come	Perceived available time for evidence review for policy drafting	- Base-line - Mid-line - End-line	Self-re-ported quantita-tive; Qua-litative	Infer-ence; Descrip-tive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
			Structures & Processes	**Opportunity (physical)**	**Utility**	Out-come	Perceived available budget for evidence review for policy drafting	- Base-line - Mid-line - End-line	Self-re-ported quantita-tive; Qua-litative	Infer-ence; Descrip-tive	T vs C ; T1 vs C; T2 vs C; T2 vs T1

				Relationships and networks	**Opportunity (social)**	**Use**	Output	# of helpdesk queries submitted by the directorate	- Midline - Endline	Monitoring data	Descriptive	
				Relationships and networks	**Opportunity (social)**	**Collaboration**	Output	Active participation to development of evidence syntheses products	- Midline - Endline	Monitoring data	Descriptive	
				Relationships and networks	**Opportunity (social)**	**Access**	Outcome	Personally knows at least one researcher affiliated to a research centre	-Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Relationships and networks	**Opportunity (social)**	**Collaboration**	Outcome	# researchers they could contact if they need support for an evidence synthesis/drafting a policy	- Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Relationships and networks	**Opportunity (social)**	**Collaboration**	Outcome	# researchers they could contact if they need support for an evidence synthesis	- Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
	RQ3	To what extent does the intervention strengthen policymakers' motivation to use evidence, operationalised as attitudes toward research, trust in research producers, and intentions to use evidence in policy processes? (Motivation)	Structured training, co-production and mentoring improves policymakers **trust** in research and researchers, as well as policy outputs Strengthening the trust in research and researchers improves the perceived **relevance and credibility** of research evidence.	**Evidence Culture**	**Motivation (perception)**	**Appraisal**	Outcome	Perceived need for research evidence for policymaking	- Baseline - Midline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence Culture	**Motivation (perception)**	**Appraisal**	Outcome	Perceived credibility of research evidence	- Baseline - Midline -	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1

			Sustained collaboration between policymakers and researchers progressively improves the **relevance of evidence syntheses** (thematically and timely)						End-line			
				Evidence Culture	**Motivation (perception)**	**Attitudes**	Outcome	Perceived competence of (domestic) research institutions	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence Culture	**Motivation (perception)**	**Appraisal**	Outcome	Perceived quality of policies (and policy outputs) when grounded in research	- Baseline - Midline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence Culture	**Motivation (perception)**	**Attitudes**	Outcome	Perceived relevance and usability of research evidence (for evidence jointly developed or supplied externally)(quality of the response)	- Baseline - Midline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence Culture	**Motivation (perception)**	**Attitudes**	Outcome	Perceived responsiveness of researchers (timely response)	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence Culture	**Motivation (intention)**	**Attitudes**	Outcome	Willingness to rely on a researcher's synthesis as a sufficient basis for a policy brief without re-running the analysis	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1

				Evidence Culture	**Motivation (intention)**	**Intention**	Outcome	Intention to use evidence in future policymaking	- Registration - Baseline - Midline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence Culture	**Motivation (intention)**	**Generation**	Outcome	Intention to generate evidence in future policymaking	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
	RQ4	Does the intervention lead to greater uptake of evidence in policy drafts, negotiations, and decision-making? (Behavioural-evidence use)	Intervention activities encourage **sustained use of research outputs** generated within study Trained and motivated participants are **consulting researchers** and i**nculcating existing evidence** within their policy discussions (and policy drafts) Trained and motivated participants use all opportunities for the **generation of evidence** with-in their work cycle	**Evidence use**	**Behaviour**	**Use**	Impact	Use of evidence repositories and archives	-Registration -Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour**	**Use**	Impact	Frequency of citing research evidence in meetings, negotiations, or internal discussions	-Registration -Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour**	**Use**	Impact	Frequency of citing research evidence in policy output	-Registration -	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1

									Baseline -Endline				
				Evidence use	**Behaviour**	**Utility**	Impact	Instances of revising or refining policy proposals based on new or existing evidence	-Registration -Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour**	**Utility**	Impact	Frequency of using evidence to rank or prioritise (trade-)policy recommendations/options.	-Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour**	**Collaboration**	Impact	Consultation of research evidence (databases, repositories, drafts, briefing notes, memos, analytical notes, RERs, evidence digests) in routine work	-Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour**	**Collaboration**	Impact	Frequency of contacting a researcher, the helpdesk, or an academic institution for input on a policy task	-Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour**	**Collaboration**	Impact	Frequency of commissioning, requesting for research evidence support on a live policy question	-Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour (institutional)**	**Institutional**	Impact	Presence of a formal evidence-review step in the	-Baseline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C;

								directorate's drafting workflow	-Endline			T2 vs T1
				Evidence use	**Behaviour (institutional)**	**Institutional**	Impact	Presence of evidence champions or focal points specialised in leveraging evidence for policy-making in the directorate	-Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour (institutional)**	**Collaboration**	Impact	Existence of partnerships with a research institute (LTA, MoU, ongoing assessment for the directorate…)	-Baseline -Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour (institutional)**	**Collaboration**	Impact	Perception that using evidence is valued, expected, and rewarded in respondent's unit.	-Baseline -Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour (institutional)**	**Institutional**	Impact	Investment of institutional time and/or financial resources into evaluations of policies	- Baseline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour (Policy)**	**Institutional**	Impact	Frequency of suggestions for policy backed by evidence/RCT	- Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour (Policy)**	**Institutional**	Impact	Perceived influence of evidence on final decision making	- Baseline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Evidence use	**Behaviour (Policy)**	**Institutional**	Impact	Personal commitment to evidence use in policy drafts,	- Baseline -	Self-reported	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C;

								negotiations, and decision-making	End-line	quantitative; Qualitative		T2 vs T1
				Evidence use	**Behaviour (Policy)**	**Institutional**	Impact	Changes in decision making practices towards evidence informed decision making	- Baseline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
Secondary	RQ5	Through **which mechanisms**, e.g., individual skill acquisition, **researcher–policymaker interaction, or organisational norms,** do interventions influence evidence uptake?		**Capability**	**Capability**	**Interpretation**	Outcome	Improvement in ability to interpret and appraise research due to training	- Midline - Endline	Qualitative	Descriptive	
				Relationships & Networks	**Behaviour**	**Use**	Impact	Extent to which interaction with researchers/mentors influenced evidence use	- Midline - Endline	Qualitative	Descriptive	
				Evidence Culture / Structures	**Opportunity**	**Use**	Outcome	Change in whether leadership, peers, and routines support evidence use	- Midline - Endline	Qualitative	Descriptive	
				Structures & Processes	**Opportunity**	**Attitudes**	Outcome	Improvement in ease of finding and using relevant research in policy workflow	- Midline - Endline	Qualitative	Descriptive	
	RQ6	How effective are trusted evidence intermediaries and embedded policy labs in bridging **the gap between research producers and policymakers** within Ethiopia's trade governance system?	Evidence intermediaries and policy lab set up increase **collaboration** (and research utility) between researchers and policymakers	**Relationships & Networks**	**Opportunity**	**Collaboration**	Outcome	Trust score for producers, policymakers, and evidence intermediaries	- Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
				Relationships & Networks	**Opportunity**	**Collaboration**	Outcome	Perception that researchers understand policy constraints and timelines	- Midline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1

				Relationships & Networks / Structures	**Opportunity**	**Utility**	Outcome	Perceived usefulness of intermediaries in translating evidence for policy use	- Midline - Endline	Self-reported quantitative; Qualitative	Inference; Descriptive	T vs C ; T1 vs C; T2 vs C; T2 vs T1
	RQ7	What are the **direct and indirect costs o**f implementing these interventions, and how do these compare wit**h their effectiveness** and potential scalability?		**Structures & Processes**	**COM**	**Utility**	Not in the ToC	Time and coordination burden of training, mentoring, and co-production activities	- Baseline - Midline - Endline	Cost benefit analysis	Descriptive	T2 vs T1
				Evidence Culture / Structures	**COM**	**Utility**	Not in the ToC	Ranking of which components produced the most useful change relative to effort	- Baseline - Midline - Endline	Cost benefit analysis	Descriptive	T2 vs T1
	RQ8	Under what conditions (institutional, political, organisational) do interventions **sustain long-term evidence use**, beyond the life of the study?		**Structures & Processes / Evidence Culture**	**Sustainability**	**Use**	Impact	Presence of leadership support, incentives, and protected time for evidence use	- Midline - Endline	Qualitative	Descriptive	T2 vs T1
				Structures & Processes	**Sustainability**	**Use**	Outcome	Existence and continued use of repositories, RER routines, and evidence processes	- Midline - Endline	Qualitative & Activity log	Descriptive	T2 vs T1
				Evidence Culture	**Sustainability**	**Utility**	Impact	Likelihood of using evidence under time pressure or political constraints	- Midline - Endline	Qualitative	Descriptive	N/A
Tertiary	RQ9	Agreement: Does co-production foster greater **consensus among policymakers** on the relevance of policy questions	Co-production improves **consensus** on evidence needs	**Relationships & Networks / Evidence Culture**	**Opportunity**	**Attitudes**	Outcome	Agreement that co-production improved consensus on priority policy questions	- Midline - Endline	Qualitative	Descriptive	

	and the types of evidence required?		**Relationships & Networks / Evidence Culture**	**Opportunity**	**Utility**	Outcome	Agreement that policymakers and researchers converged on appropriate evidence types	- Midline - Endline	Qualitative	Descriptive	
RQ10	Collaboration: Does the intervention **strengthen cross-ministerial collaboration** in evidence use, particularly through joint training, shared RERs, and policy dialogues?	Joint training and other shared research outputs improve **collaboration** between policymakers	**Relationships & Networks**	**Opportunity**	**Collaboration**	Outcome	Frequency of cross-ministerial exchanges on evidence and policy questions	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
			Relationships & Networks / Structures	**Opportunity**	**Collaboration**	Outcome	Contribution to shared briefs, RERs, or analytical outputs	- Baseline - Midline - Endline	Self-reported quantitative	Inference	T vs C ; T1 vs C; T2 vs C; T2 vs T1
RQ11	Equity and inclusion: Do interventions affect female policymakers differently in terms of capability, opportunity, or motivation to use evidence? (important for gender balance and subgroup analysis).		**Capability**	**Capability**	**Interpretation**	Outcome	Gender differences in self-reported analytical capability and confidence	- Baseline - Midline - Endline	Self-reported quantitative	Descriptive	Male vs female
			Relationships & Networks / Structures	**Equity in Evidence Use**	**Collaboration**	Outcome	Gender differences in access to mentoring, networks, and resources	- Midline - Endline	Self-reported quantitative	Descriptive	Male vs female
			Evidence Culture / Behaviour	**Behaviour**	**Use**	Impact	Gender differences in self-reported evidence use in policy tasks	- Baseline - Midline - Endline	Self-reported quantitative	Descriptive	Male vs female

Source: Own elaboration

Annex 3c: Stakeholder Engagement and Evidence Uptake Pla

1. Purpose

The success of this project, in its implementation and outcome, i.e. evidence uptake in policy making, largely depends on the level of collaboration and co-creation of evidence with core stakeholders throughout the process. Building and sustaining collaboration requires transparent communication, inclusive engagement, and the application of adaptive management practices. Given the institutional nature of the intervention, high-level participation and sustained support from key stakeholders are essential to achieving the project's objectives. Accordingly, the Stakeholder Engagement and Evidence Uptake Plan (SEEP) is designed to increase evidence uptake through actively engaging stakeholders throughout the all the project phases.

Systematically strengthening stakeholders' participation and support across all stages of the project is the primary purpose of this SEEP. Strengthening institutional relationships will not only improve collaboration but also increase commitment to own the process and the outcomes of the project. The level and form of engagement will depend on project milestones and stakeholders' categories. Early-stage engagement is particularly critical, as it lays the foundation for subsequent phases of implementation. In the initial phase, especially during the identification and selection of participants from each institution, senior leadership from the respective line ministries and commissions will be directly engaged to ensure institutional alignment, ownership, and high-level endorsement. Such early involvement is essential for securing commitment and facilitating smoother implementation in later stages.

Overall, this SEEP seeks to achieve the following four key objectives:

1. **Enhance Participation and Commitment of Institutions:** Promote active and informed participation of all participating institutions and relevant governmental stakeholders at each stage of the project.
2. **Promote Co-ownership of the Processes and Outcomes:** Designed to ensure that participating institutions co-own both the intervention process and its outcomes. Through joint planning, collaborative evidence generation, and participatory validation of findings, the project will embed institutional responsibility and increase the likelihood of sustainability and post-project continuity.
3. **Mitigate Operational, Institutional and Political Risks:** The SEEP aims to systematically identify and manage risks that may affect project implementation, study integrity, or policy relevance. Through maintaining continuous engagement and feedback loops with key stakeholders, the project will employ adaptive management approaches to respond to emerging institutional, administrative, or political challenges.
4. **Promote Evidence-Based Decision Making through Stakeholder Engagement:** Strengthen the capacity of stakeholders to interpret and apply evidence and establish feedback mechanisms to monitor how research informs policy decisions.

2. Evidence uptake and use objectives

National, sub-national or regional influence (e.g., policy decision-makers, evidence broker organisations, researchers, national or regional networks)

Objectives	*Indicators*
Strengthening policymakers' technical capacity and practical use of research evidence in policy processes, as well as improving their access to high-quality research (national)	• *Policymakers demonstrate improved technical skills (policy analysis, data interpretation, evidence appraisal) and use of research evidence. These are to be measured through baseline-endline scores, and performance in Rapid evidence review (RER) exercises.* • *Increased frequency and quality of evidence use in policy documents (memos, briefs, technical notes) and decision-making processes. This will be verified through document reviews as well through baseline-endline responses.* • *Number of policy briefs, technical notes, or analytical reports produced by trained policymaker during and after the training phase of the intervention.*
Strengthen sustained collaboration and exchange between researchers and policymakers (national)	• *Number of co-created research outputs (Phase II co-creation phase).* • *Number and functionality of mentor groups formed and engaged.* • *Invitations to participate in formal or informal working groups on policies or programmes that are related to the research focus.* • *Instances of stakeholder feedback demonstrating engagement with research outputs (e.g. comments, revisions, clarifications, follow-up analysis requests)*
Foster an organisational culture where evidence use becomes routine, credible, and sustainable (national)	• *Policy documents, strategies, or briefing notes that incorporate quality research evidence.* • *Citations of the research findings in internal or external policy documents as justification for action.* • *Instances where findings have been used to influence policy agendas or inform decisions of key stakeholders (e.g. meeting minutes, policymaker reporting at endline).* • *Number of downloads, log-in metrics of research outputs, including policy briefs, and usage of research databases such as JSTOR, World Bank eLibrary, ScienceDirect, and RePEc). This will be tracked user ID (on the repository) and through reported use in endline survey.*

	• *Changes in institutional practices of expectations regarding evidence use (e.g. policy guidelines, routines, requirements to reference evidence).*

Wider evidence ecosystem influence (e.g., multilaterals, funders, international networks, global evidence infrastructure)	
Objectives	*Indicators*
Generate operational evidence on how capacity-building interventions strengthen evidence uptake and use in policymaking processes to inform donors and multilateral organisations on EIPM (FCDO, WB, AfDB).	• *Evidence that the findings of interventions are informing other capacity-building programmes in Ethiopia or beyond. It will be measured through official requests from other organisations to use the project's materials or approach.* • *Invitations to (and incidences of) present findings and disseminate to multilateral organisations (e.g. FCDO, World Bank, AfDB).*
Increase engagement of policymakers with global and regional evidence networks and knowledge platforms (Africa Evidence Network)	• *Policymakers actively engage with global evidence infrastructure or international research networks for accessing policy-relevant evidence.* • *Participation rates in stakeholder learning events (workshops, roundtables) involving international and regional actors.* • *National policy documents, strategies, or technical reports that reference or incorporate international network research evidence.* • *Number of downloads or views of project outputs (reports, training materials, RERs) by external audience.*

Evidence-use research frontier (e.g., filling gaps in the EIPM evidence base, furthering academic understanding, methodological innovation)	
Objectives	*Indicators*
Introduce rigorous causal approaches (RCTs and other high rigour designs) to evaluate interventions that are rarely tested with rigour in governance settings.	• *Invitations to present the findings at EIPM conferences or methods workshops.*

	• *Invitations to write methodological chapters or synthesis work (including EIPM syntheses).* • *Publication of working papers, policy reports, or journal submissions emerging from the study.* • *Number of citations of the study findings in other (EIPM) academic studies and journals.* • *Adoption of project analytical framework by other national and international researchers (in their studies or frameworks).*
Address critical evidence gaps in literature by examining how awareness, consensus on evidence priorities, and relational trust influence the uptake and use of research evidence in decision-making processes.	• *Number of citations of the findings in academic journals.* • *Evidence that the study analytical framework (awareness, consensus, relational trust) is being used in subsequent EIPM systematic reviews or research related to EIPM. Measured through download and dissemination metrics of methodological or analytical outputs (on google scholar).* • *Evidence that capacity-building practitioners are using your findings to design their programmes.* • *Initiation of programmes or pilot interventions in response to research findings, in the same or other contexts.* • *Evidence that subsequent EIPM studies build on this study theoretical and qualitative insights.*

3. Key stakeholders

a. Please describe how this project came about and the extent to which the project responds to specific demand.
Ethiopia's trade governance has evolved from a developmental state model, combining tariff reductions, simplified licensing/foreign exchange allocation, export-promotion zones, and industrial parks under strong regulatory oversight, to a more market-oriented, private-sector-led approach. Ethiopia is implementing AfCFTA commitments and engages in COMESA and globally, it has resumed WTO accession, with negotiations re-energized since 2020 and a target of accession by 2026. Estimates indicate that WTO membership alone could raise real GDP by 8–10%. Realising these benefits depends on evidence-informed policymaking that guides negotiations, domestic reform sequencing, and institutional alignment.

This project identifies three main groups of stakeholders: core (primary) stakeholders, strategic (secondary) stakeholders, and external stakeholders. Each group plays a distinct but complementary role in supporting the design, implementation, oversight, and sustainability of the project. The specific institutions and actors under each category are outlined below.

Core/Primary Stakeholders: This core group comprises ministries and commissions that are directly engaged in the implementation of the project and serve as key partners in its execution. Primary stakeholders include the Ministries of Trade and Regional Integration; Labour and Skills; Agriculture; Industry; Finance; and Planning and Development, as well as the Ministry of Revenue. It also includes the Ethiopian Customs Commission, the Ethiopian Commodity Exchange, and the Ethiopian Investment Commission. These institutions play a central role in delivering project activities and achieving intended outcomes.

Strategic/Secondary Stakeholders: These stakeholders play a critical oversight and approval role in policymaking, supervise policy implementation. This includes the Office of the Prime Minister, the Prosperity Party, and the Parliament Standing Committee on Trade and Industrial Development. As strategic actors within the policymaking ecosystem, they provide the institutional framework within which evidence-informed policymaking can be reinforced and sustained.

External Stakeholders: This includes donors, consultants and private sector. Accordingly, we have identified the following institutions as external stakeholders, these are Business membership organisations, FCDO and Policy Studies Institute of Ethiopia.

In each institution, this plan target directorates, departments, and technical units that play direct roles in policy formulation, strategic planning, regulatory oversight, programme implementation, data production, and analytical research. These include, but are not limited to, policy analysis units, planning and monitoring divisions, research and development departments/units. The engagement efforts focus on functionally strategic units rather than entire institutions. At the same time, the approach facilitates coordinated engagement across interconnected policy domains, enabling the project to address systemic bottlenecks that cut across ministerial boundaries.

b. Have policy counterparts been involved in the design of this study?

☐ *Yes, the study was co-designed with policy counterparts (By 'co-design' we mean a collaborative and participatory process in which policy counterparts and researchers work together to identify evidence needs and the means to address them).*

☐ *Yes, we sought feedback from policy counterparts on the research design.*

☒ *No, but we plan to engage policy counterparts during the inception phase of the project.*

☐ *No, we have no immediate plans to engage policy counterparts in the design of the study.*

If Yes, please provide details. If No, please expand.

During the inception and baseline workshop, policymakers will play an active role in identifying sector-specific evidence needs and co-designing the research agenda and evidence production process. This collaborative approach ensures that research priorities, methodologies, and outputs are directly aligned with policy demands, strengthening ownership and increasing the likelihood of evidence uptake.

c. *Identify at least three stakeholders with the potential to influence the uptake and use of research findings and explain your selection* *(we would expect there to be more than three in most instances). We encourage you to consider a range of key stakeholders who exercise power to influence and make change happen in different ways, formally and informally, in the community, government, parliament, the media, civil society, the private sector, academia, religious organisations, traditional leadership, trade unions, professional associations, etc. Insert more rows as needed.*

Table 1: Level of Engagement of Stakeholders

Name of key stakeholder	Organisation name and website	Relevance or level of influence	Previous experience of work with stakeholders
	Primary Stakeholders[5]		
Jemal, Team Leader of Statistics and information	Ministry of Trade and Regional Integration	High/Collaborator	EA has organised several regional and national policy dialogue forums in collaboration with the ministry.
Eshetu, Management CEO	Ministry of Industry	High/Collaborator	EEA in collaboration with the ministry, provided training to each regional Chamber and commerce associations.
Dr Desta, Executive Director of Policy studies and Research	Ministry of Labour and Skills	High/Collaborator	As this ministry was established recently, we have not yet had the opportunity to collaborate.
Tesfaye, Director of Policy Research	Ministry of Agriculture	High/Collaborator	Over the past two years, EEA collaborated with the Agricultural Transformation Institute on two projects (ATI)[6].

[5] EEA organizes an annual International Conference, which is now in its 23rd edition. The ministries have actively participated in and contributed to this conference over the past twenty years.

[6] Institutional Frameworks and Structural Set up for Pluralistic Agricultural Extension and Advisory Service (PAEAS) and Institutional, Policy and Regulatory Analysis to Enabling the Business of Agriculture in Ethiopia.

Abiti, Director of Training and HR	Ministry of Finance	High/Collaborator	EEA has signed MoU agreement to work with MoF on policy relevant issues.
Habtamu, Director of Development Planning	Ministry of planning and Development	Moderate/Collaborator	EEA conducted a joint research project titled "Rebuilding, Restitution, Rehabilitation, and Reconstruction of the Post-Conflict Ethiopian Economy" in 2022.
Ousman, Director of Strategic Partnership	Ministry of Revenue	High/Collaborator	EEA has worked several projects in collaboration with regional revenue Bureaus (Afar, SNNPRS, Somali and Amhara).
Abdisa, director planning and budget	Customs Commission	High/Collaborator	No joint research projects have been undertaken.
Habtamu, Team Leader of Education and training	Investment Commission	Moderate/Collaborator	EEA conducted a research study titled "Private Sector Development in Ethiopia: Trends, Challenges, and Policy Issues" In 2022.
Bayelign, Strategic Plan and Research Officer	Commodity Exchange	Moderate /Collaborator	No joint research projects have been undertaken.
	Strategic/Secondary Stakeholders[7]		
	Office of the Prime Minister	Moderate	Former EEA board members serve as policy advisors to the Prime Minister's office.
	Civil Service Commission	Low	No joint research projects have been undertaken.

[7] The contact details of each secondary and external stakeholder will be incorporated into the Stakeholder Engagement and Evidence Uptake Plan after the inception workshop. The EEA will invite the secondary and external stakeholders to participate in the inception workshop.

	Parliament (Standing committees)	Moderate	EEA is a member of the Parliamentary Research Network.
	Prosperity Party	High	No joint research projects have been undertaken.
	External Stakeholders		
Esteban Quinones	FCDO	High/Partner	No joint research projects have been undertaken.
	Business Membership Organisations	Low	Partnered with the association to provide training for member businesses.
Gashaw Abate and Kibrom Abey	IFPRI	Low	Various EEA members are associated with organisation
Haleform Nigus	Policy Studies Institute	Low	Collaborated in joint workshops and research evidence generation.

4. Engagement and communication plan

Please use the table below to plan activities and products through your research cycle*. You will be expected to report your progress on the activities identified here as part of your tranche-linked progress reporting. Please provide supporting documents, including participant lists, key takeaways and follow-up plans from meetings and events. We encourage you to share copies of presentations, blogs, articles, policy briefs, memos and other related knowledge products, including publications related to this research study. Add rows to the table as necessary.*

Stakeholders will actively participate throughout the key phases of the project. The specific nature, timing, and methods of engagement for each stakeholder group are detailed in Table 2, which outlines the comprehensive engagement plan.

The participants are civil servants who are often required to attend multiple donor-funded events. To accommodate their schedules, the programme is designed to include three consecutive days of in-person training, followed by an eight-week blended training and mentorship component delivered once per week (half-day or full-day sessions). In total, participants will be required to commit a maximum of 12 days including the inception workshop in full engagement over a 24-month period, which is considered reasonable given their existing responsibilities. In addition to these structured sessions, participants will continue contributing to evidence co-creation activities as part of their regular office duties.

1. **Planning:** The detailed timeline for how the capacity building and technical support will unfold, highlighting important dates, delegating tasks, and available resources documents will be shared with primary stakeholders before the commencement of the project.
2. **Execution**: The stakeholders will engage in this phase through assigning focal person to collaborate with the research team in identifying the problem for evidence generation and documenting the lessons learning sessions through the presentation of synthesised evidence on issues related to each organisation and communicate with the rest of the team in the organisation.
3. **Monitor:** Quarterly update/report on the progress on capacity building activities including mentorship and the progress of co-production of evidence will be briefed to the stakeholders to track the progress of the project and to make adjustment, if required.
4. **Closure/evaluation: Evaluation of the project in terms of participation, trust and collaboration and insight into further improvement in the future.**

Table 2: Stakeholder Engagement Plan[8]

Key milestones and timeline	Why- objectives of the engagement	Who- stakeholders	How- channels	When- timing	By whom
Formal Mobilisation & Institutional Commitment	To raise awareness and secure institutional buy-in	Primary	Official Communication	Early March '26	Prof. Mengistu Ketema
Participant Identification & Validation	To secure ownership	Primary	Jointly review and validate the profiles of participants	End March '26	Dr Mezid Nasir
Workshop	Baseline information	Primary	Participant registration and post randomisation invitation and participation in workshop	June '26	Dr. Abule Mehare
Implementation	Capacity building	Primary	Training, mentorship, technical sessions	June-September '26	Dr. Atika Pasha

[8] This plan will be reviewed and updated as necessary throughout the project.

	Evidence culture is developed across and within ministries	Primary	Evidence co-production and mentoring	September '26- May '27	Dr. Atika Pasha
Monitoring & Adaptive Management	Progress update	Primary	Review meetings with senior leadership	June '26; Aug '27	Dr Johanna Gather & Dr. Abule Mehare
Results Validation & Knowledge Sharing	Disseminating research findings	All groups of Stakeholders	Validation workshops	Dec '27	Prof Markus Frölich and Dr. Abule Mehare

5. Why will stakeholders want to engage?

The stakeholders are senior and mid-level civil servants who operate within demanding institutional environments characterised by competing priorities and frequent participation in externally funded programmes. Therefore, engagement under this project is designed to offer clear and professionally relevant incentives that justify their time commitment.

First, participation provides structured professional development opportunities. The planned capacity building training sessions in econometrics, causal inference, evidence synthesis, and statistical reasoning will strengthen their analytical, policy evaluation, and evidence-use capacities through targeted training and hands-on mentorship. This contributes directly to enhancing their effectiveness in policy formulation, implementation, and review processes.

Second, the project offers access to research resources and knowledge infrastructure, including curated research databases, policy-relevant literature such as (JSTOR, World Bank eLibrary, ScienceDirect, RePEc). Such access supports informed decision-making and improves the quality of policy outputs within their respective institutions.

Third, stakeholders will benefit from structured mentorship and peer-learning mechanisms embedded within the project design. Mentor–mentee groups will be established, pairing researchers with participating stakeholders to facilitate continuous technical guidance and practical learning. This will strengthen individual competencies of the participants and promote institutional learning.

Fourth, active participants will have the opportunity for recognition and intellectual contribution, including co-authorship of policy briefs, and knowledge products developed through the evidence co-creation process. This enhances professional visibility and supports career advancement.

Finally, through co-creation of evidence, engagement ensures that primary stakeholders have direct input into shaping research agendas and policy outputs relevant to their mandates. This alignment increases the practical relevance of project deliverables and supports institutional performance objectives.

6. Monitoring and review

The engagement process will focus on consolidating learning, capturing feedback, and promoting policy uptake. This will involve:

- Structured feedback tools and post-engagement surveys to measure the participation rates in meetings and workshops
- Synthesis of lessons learned and cross-institutional insights

Research findings and stakeholder inputs will be communicated back to institutions through policy briefs, presentations, and technical notes. This ensures that institutional contributions are reflected in final outputs and that evidence generated through the process informs ongoing policy debates and future reform initiatives.

7. Risk and mitigation measures

The potential risks, their qualitative likelihood of occurrence, and the corresponding mitigation measures are summarised in Table 3. Based on the preliminary risk assessment, five key risks have been identified. The first three risks are directly related to project implementation and stakeholder engagement. These include limited stakeholder participation in capacity-building activities, insufficient engagement in learning sessions and evidence co-creation processes, and low uptake of project findings and recommendations. These risks may affect the effectiveness, credibility, and practical relevance of the intervention if not proactively managed. The remaining two risks are contextual and relate to the broader institutional operating environment. These include turnover of senior leadership and limited institutional capacity to absorb and sustain the outcomes of the intervention. Such risks may influence continuity, ownership, and long-term sustainability beyond the project period.

Mitigation measures are therefore designed to strengthen institutional commitment, ensure continuous engagement across different levels of leadership and technical staff, embed adaptive management practices, and promote integration of project outputs into existing institutional systems and processes. Together, these measures aim to safeguard implementation quality while enhancing sustainability and policy relevance.

Table 3: Potential Risks and Mitigation Measures

Risk	Description	Probability	Mitigation Measures
Limited Stakeholder Availability	Competing institutional priorities may limit participation in trainings, meetings, or review sessions.	Moderate	• Secure early high-level endorsement to prioritise engagement. • Schedule activities in consultation with institutions. • Provide advance notice and flexible formats (virtual/hybrid)

Limited engagement in evidence co-creation	Stakeholders may not actively contribute to data validation, interpretation, or joint learning processes.	Low	• Use participatory workshops and structured feedback tools. • Clarify benefits of co-creation for institutional decision-making. • Assign institutional focal people to coordinate input
Low finding uptake	Institutions may not integrate lessons or recommendations into policy or practice.	High	• Engage leadership early in validation of findings. • Align recommendations with institutional mandates and priorities. • Develop practical action plans jointly with institutions
Leadership turnover	Changes in senior officials may disrupt commitment and continuity	Moderate	• Institutionalise engagement beyond individuals. • Maintain formal documentation and official correspondence. • Conduct periodic briefings for new leadership.
Limited institutional capacity to absorb training	Skills gained may not translate into practice due to systemic constraints	Moderate	• Combining training with mentoring and follow-up coaching. • Aligning activities with institutional workflows

8. Inclusivity markers

Please use this section to consider the extent to which your project addresses issues related to gender or marginalised or vulnerable groups.

a. The following statements relate to the extent to which your project addresses *gender issues*. Please select the <u>one option</u> that most closely aligns with your project.
☐ Addressing issues related to gender is the primary focus of my project. For example, select this option if your project explicitly addresses an issue pertinent to women/girls, uses a gender informed design and/or analyses gender separate effects to consider recommendations specific to gender.

☒ Addressing issues related to gender is a secondary or parallel focus of my project. *For example, select this option if your project targets a broader group than women/girls but intends to provide insights based on gender separate effects or gender more broadly, to answer the research question or inform research objectives.*

☐ My project does not have an explicit focus on gender

☐ Unsure - *it is too early to tell with certainty whether my project will include any explicit focus on gender issues.*

If relevant, please provide detail on the way in which your project addresses issues related to gender, in particular identifying whether this will involve data collection and/or analysis.

Gender inclusion is an important consideration in this project. The project will promote equitable participation and aim to ensure that women represent 33% of participants in project activities, including training, workshops, and engagement platforms.

b. The following statements relate to the extent to which your project addresses issues faced by *marginalised or vulnerable groups*. Please select the one option which most closely aligns with your project. *Depending on the context of your research, marginalised or vulnerable groups could, for example, include: people living with disability, geographically and politically marginalised groups, and youth and elderly groups*

☐ Inclusivity of marginalised or vulnerable groups is the primary focus of my project. *For example, select this option if your project explicitly addresses an issue pertinent to marginalised or vulnerable group(s) or seeks to find and analyse separate effects on marginalised groups with the aim to provide recommendations or insights to improve inclusion.*

☐ Inclusivity of marginalised or vulnerable groups is a secondary or parallel focus of my project. For example, select this option if your project targets a broader group beyond those who are marginalised or vulnerable, but intends to provide insights into differential effects on these groups to answer the broader research question or inform the broader research objective.

☒ My project does not have an explicit focus on marginalised or vulnerable groups.

☐ Unsure – *it is too early to tell with certainty whether my project will include any explicit focus on marginalised or vulnerable groups.*
If relevant, please provide detail on the way in which your project addresses issues faced by specific marginalised and vulnerable group(s), in particular identifying whether this will involve data collection and/or analysis.

Annex 3d: Added robustness checks

Table 4. Three level DEFF sensitivity calculations (using Powerup on Excel)

Scenario	N ind (HH per directorate)	ICC1 (directorate)	ICC2 (ministry) - Low	ICC2 (ministry) - Mid	ICC2 (ministry) - High	DEFF (Low ICC2)	DEFF (Mid ICC2)	DEFF (High ICC2)	Effective N endline (Low)	Effective N endline (Mid)	Effective N endline (High)
70 clusters, T=42, C=28, ICC1=0.1, n_ind=5	5	0.10	0.05	0.10	0.15	2.650	3.900	5.150	95	65	49
60 clusters, T=36, C=24, ICC1=0.1, n_ind=5	5	0.10	0.05	0.10	0.15	2.650	3.900	5.150	82	55	42
60 clusters, higher T attrition, ICC1=0.1, n_ind=5	5	0.10	0.05	0.10	0.15	2.650	3.900	5.150	82	55	42
60 clusters, ICC1=0.25, n_ind=5	5	0.25	0.05	0.10	0.15	3.250	4.500	5.750	66	48	38

Source: Own calculations