GREEN CLIMATE FUND | Independent Evaluation Unit

IFAD

EVIDENCE REVIEW ON BEHAVIOURAL SCIENCE INTERVENTIONS IN DEVELOPMENT AND ENVIRONMENTAL FIELDS IN DEVELOPING COUNTRIES

Protocol

Samantha Booth, Romina Cavatassi, Benjamin Curtis, Deborah Sun Kim, Yeonji Kim, Laurenz Langer, Caitlin Blaser Mapitsa, Elangtlhoko Mokgano, Promise Nduku, Martin Prowse, Jyotsna Puri, Jamie Robertsen

June 2022

GREEN CLIMATE FUND
INDEPENDENT EVALUATION UNIT

INTERNATIONAL FUND FOR
AGRICULTURAL DEVELOPMENT

Evidence review on behavioural science interventions in development and environmental fields in developing countries

PROTOCOL

Samantha Booth, Romina Cavatassi, Benjamin Curtis, Deborah Sun Kim, Yeonji Kim, Laurenz Langer, Caitlin Blaser Mapitsa, Elangtlhoko Mokgano, Promise Nduku, Martin Prowse, Jyotsna Puri, Jamie Robertsen

06/2022

CONTENTS

TABLES

ACKNOWLEDGEMENTS

The authors are grateful for the comments received from the engagement committee for this protocol. The engagement committee includes Chaning Jang (Busara), Nathanial Peterson (Busara), Erik Thulin (Rare), and Katie Williamson (Rare). Any errors or inconsistencies are entirely the responsibility of the authors.

LIST OF AUTHORS

The authors of the protocol of the evidence review on behavioural change in developing countries are (in alphabetical order by last name):

FULL NAME	AFFILIATION
Samantha Booth	Human Friendly Solutions
Romina Cavatassi	International Fund for Agricultural Development
Benjamin Curtis	Human Friendly Solutions
Deborah Sun Kim	Independent Evaluation Unit, Green Climate Fund
Yeonji Kim	Independent Evaluation Unit, Green Climate Fund
Laurenz Langer	Africa Centre for Evidence, University of Johannesburg
Caitlin Blaser Mapitsa	School of Governance, University of the Witwatersrand
Elangtlhoko Mokgano	Independent Evaluation Unit, Green Climate Fund
Promise Nduku	Africa Centre for Evidence, University of Johannesburg
Martin Prowse	Independent Evaluation Unit, Green Climate Fund
Jyotsna Puri	International Fund for Agricultural Development
Jamie Robertsen	Independent Monitoring, Evaluation and Learning Consultant

A. BACKGROUND

1. DESCRIPTION OF THE PROBLEM

Climatic change is projected to intensify over the next several decades, resulting in dramatic impacts on natural and human systems. The need for both adaptation and mitigation is clear. It is equally clear that human behaviour is a key driver of climate change; therefore, many adaptation and mitigation strategies require changes in behaviour. There is currently a lack of rigorous empirical evidence on what could encourage or potentially bring about a change in human behaviour in a way that would be useful for reducing greenhouse gas emissions and adapting to the changed climate, especially in developing countries that are more vulnerable to the impacts of climate change. This is a pressing problem given that climate change's impacts will occur unevenly across the globe and disproportionately affect developing countries due, in part, to their limited capacity to deal with shocks, stresses and damaging fluctuations (see Global Commission on Adaptation, 2019; Solomon and others, 2007; Intergovernmental Panel on Climate Change, 2014; United Nations Environment Programme, 2017; Wade and Jennings, 2015; Binet and others, 2021). Burning fossil fuels and other anthropogenic activities are the primary drivers of climate change. Transportation, energy consumption and production, and food production present some of the most significant opportunities to change human decisions and activities to reduce carbon emissions (Williamson and others, 2018).

In recent decades, theories and evidence from behavioural science – defined by Balmford and others (2021) as the scientific study of behaviour informed by an array of disciplines, including sociology, psychology, economics, anthropology, and political science – have provided insights into the social, motivational, cognitive, cultural, and contextual factors underlying human behaviour. Stern (2020) describes behavioural science interventions as involving neither command and control regulations nor solely financial incentives to change behaviour. Examples include information provisions, appeals to values and norms, or engagement and restructuring choice options. These insights have informed interventions that have helped to encourage socially valued behaviour change, including reductions in smoking, addiction, and obesity as well as improvements in tax compliance, development assistance, and climate change mitigation (Duflo and others, 2011; Datta and Mullainathan, 2014; Hallsworth and others, 2017; Bollinger and others, 2020). Research has informed behaviour change interventions relevant to various environmental issues, including, but not limited to, energy efficiency, water conservation, recycling, and transport (Osbaldiston and Schott, 2012; Byerly and others, 2018; Nisa and others, 2019).

There is an opportunity and a responsibility to affect change through increased understanding of the factors underlying the anthropogenic causes of climate change and ways that mitigation and adaptation behaviours may be effectively encouraged (Gifford and others, 2011). Insights from behavioural science have been frequently applied to enhance public policy effectiveness (OECD, 2017). For example, nudges as a category of psychology-based interventions can be a cost-effective tool to support individual decision-making and have been applied to foster pro-environmental behaviours (Cinner, 2018; Schubert, 2017). Nudges can involve simple alterations to the physical microenvironments in which choices are made (choice architecture). Such small changes can significantly affect behaviour, helping people make decisions that benefit themselves and the broader community (Szaszi and others, 2018; Thaler and Sunstein, 2009; Hollands and others, 2017). Against this background, this protocol for a systematic review focuses on synthesizing the evidence on the effectiveness of behavioural science interventions, including feedback, reminders, salience (communication), salience (experience design), and goal setting in promoting environmental and development goals in developing countries. This protocol presents the overall approach for the systematic review of the effectiveness of these specific behavioural science

interventions on environmental and development outcomes in developing countries, with a particular focus on the data collection and analysis.

2. WHY IS IT IMPORTANT TO DO THIS REVIEW?

This review examines the effectiveness of specific behavioural science interventions. They include feedback, reminders, salience and goal setting in promoting environmental and development outcomes by individuals, households, communities and companies in developing countries. The review's focus has its origins in the growing confidence in behavioural science interventions (Schott and others, 2016; Stern and others, 2016) as potentially cost-effective strategies compared to traditional market tools and regulation. To the best of our knowledge, there is no systematic review evidence that carefully explores the effectiveness of these behavioural science interventions on environmental and development outcomes in developing countries. Extensive evidence exists about what works and what does not in promoting behaviour change broadly (Flanagan and Tanner, 2016). But this evidence base has not been rigorously synthesized in relation to climate change in developing countries. This review reduces this gap within the literature to inform the Green Climate Fund, the International Fund for Agricultural Development, broader multilateral agencies, development practitioners and other decision makers about the available evidence on a broad set of behavioural science interventions in developing country contexts and the extent to which they contribute to desired environmental and development outcomes. Appendix 1 describes the interventions in detail and Table 1 offers a summary of each of the intervention types.

Table 1. Behavioural science intervention definitions - feedback, reminders, salience and goal setting

BEHAVIOURAL INTERVENTIONS		DEFINITION
WHEN is the choice made? This category of interventions encourages positive choices by influencing key decisions.	Reminders	This type of intervention involves messaging people (via email, text message, etc.) in a timely way to call their attention to something and/or to encourage them to take certain actions.
	Feedback	This type of intervention provides information, often tracked over time, about behaviours. The information might report how the tracked behaviours compare to targets and/or outline consequences of the behaviour trajectories.
WHICH choices are available? This category of interventions encourages positive choices by altering the set of options available.	Salience (communication)	This type of intervention improves the ease and accessibility of adopting behaviours by making information/choices more prominent and relevant. Personalising communication and highlighting follow-on instructions are typical strategies to increase salience. Because this intervention focuses on messaging content rather than timely delivery, it is distinct from a reminder.
	Salience (experience design)	This type of intervention targets how individuals interact with their physical and/or digital environment. It involves arranging facilities or options so that they are either: (1) more prominent, accessible, and easy to prompt a particular behaviour or, (2) less prominent, accessible, or easy to discourage a particular behaviour.
	Goal setting	This type of intervention helps individuals consider what their priorities are, then specify a series of goals that they would like to achieve. It often goes along with a planning process.

B. Objectives of the review

The primary objective of this protocol for a systematic review is to identify, assess and synthesize evidence on the effectiveness of feedback, reminders, salience and goal setting interventions conducted in developing countries on environmental and development outcomes. It facilitates the use of evidence in informing policy and practice within the environmental and development fields, particularly climate mitigation and adaptation. In doing so, we address the following review questions:

- What is the impact and effectiveness of feedback, reminders, salience (communication), salience (experience design) and goal setting on environmental and development outcomes?
- To what extent do effects vary by population characteristics, evaluation design, intervention type and time period after the intervention?
- To what extent do implementation features moderate the effectiveness of these behavioural science intervention programmes?

C. Methods

1. The overall systematic review design approach

We use a two-stage evidence review approach. The first stage consists of a completed evidence gap map (EGM).[1] The second stage consists of conducting a systematic review and synthesis in compliance with the Campbell Collaboration's guidelines for producing systematic reviews (SRs).[2] We adopt an effective and adaptable research process that fully integrates the selection of cells for the systematic review (SR) from the completed EGM. Previous synthesis projects in the environmental sector (see Snilstveit and others, 2019; Langer and others, 2018) indicated the successful integration of an evidence map and subsequent full systematic review is dependent on four key factors:

- Continued and embedded stakeholder engagement on the scope of the overall project and synthesis outputs
- A consistently rigorous and transparent synthesis approach that applies similar criteria to both outputs (the EGM and the SR)
- A sufficiently broad scope and design of the EGM that guarantees a sufficient evidence base for subsequent synthesis
- A versatile software solution to provide flexibility in the evidence mapping tool to integrate the knowledge management aspect of the evidence review with the visualization requirements of the EGM

a. Evidence gap map

The EGM's inclusion of evidence had a broader scope than the full systematic review. But both are focused on the nature of existing evidence regarding the effectiveness of behavioural science interventions on environmental and development outcomes in developing countries. The EGM

[1]Details on the theory of change, intervention-outcome framework, inclusion/exclusion criteria, the search strategy, screening and data management are provided in the approach paper and EGM report. See *https://ieu.greenclimate.fund/evidence-review/behavioural-science*.

[2] For systematic reviews *https://onlinelibrary.wiley.com/pb-assets/Campbell%20Policies%20and%20Guidelines%20Dec2020-1608292090217.pdf* and for evidence and gap maps *https://onlinelibrary.wiley.com/doi/10.1002/cl2.1125*.

mapped evidence from impact evaluations and SRs across 22 behavioural science interventions. Its main objectives were to indicate the overall nature and size of the available evidence base, identify areas for synthesis, and substantiate evidence gaps for future analysis. The systematic review will focus on five of the 22 selected interventions: feedback, reminders, salience (communication), salience (experience design), and goal setting.

The EGM's evidence base supports stakeholder engagement and decision-making about the most effective synthesis approach and scope. The final map has 84 studies (82 impact evaluations and two SRs). The EGM guided discussions about which areas of the evidence base to use for synthesis and the most effective method for synthesizing the evidence to answer the review question. Following two meetings with the advisory group to identify the relevant areas of evidence for the systematic review, both stakeholder interest and a sufficient body of evidence for specific cells in the EGM steered the review's focus towards five interventions: feedback, reminders, salience (communication), salience (experience design), and goal setting interventions.

b. Systematic review and synthesis

Sixty-eight unique studies from the EGM were identified as focusing on these five intervention categories. An effectiveness analysis will be conducted to answer the review questions regarding the effectiveness of these interventions in achieving behavioural change in selected environmental and development outcome areas in developing countries. Therefore, the systematic review will only include primary studies that measure the effects of interventions and whose design can reliably attribute observed effects to these applied interventions. Individual effects will be synthesized into overall estimates of treatment effects using statistical meta-analysis.

2. Theory of change

In the context of the evidence review on behavioural science interventions, the purpose of the theory of change is to inform the types of interventions included in the systematic review. An extensive description of the theory of change may be found in the approach paper and the EGM report. The theory of change also helps to highlight possible moderators in the meta-analysis.

3. Intervention-outcome framework for the EGM

The EGM intervention-outcome framework is the primary tool used to structure and visualize the evidence base. The theory of change directly influences its design. The approach paper and EGM report provide comprehensive details of the intervention-outcome framework.

4. Criteria for inclusion and exclusion of studies in the review

To systematically synthesize literature on the effectiveness of the selected behavioural science interventions, an underlying focus on environmental and human development outcomes guides the scope of the review. We use the PICOS (Population, Intervention, Comparator, Outcome and Study design) framework to develop the inclusion criteria. The approach paper and EGM report contain full details of the systematic review's inclusion criteria. The inclusion criteria define the precise characteristics of studies that will be included in the systematic review. All evidence not meeting these criteria will be excluded from this review.

5. Searching for evidence

We developed a comprehensive search strategy to identify qualifying studies and all available evidence relevant to the review question for inclusion in the systematic review. The approach paper

and EGM report outline the search strategy, including sources (databases and repositories), backward and forward reference searches, combination of search terms, and results from the searching and screening process.

6. DATA COLLECTION AND ANALYSIS

a. Data extraction and management

We will use a predefined data extraction tool to extract data systematically and transparently from the included primary studies. The coding tool highlighted in Appendix 2 will be transferred into Excel to extract information necessary for the systematic review and synthesis. Full-text reports will be examined and studies coded on variables related to:

- Descriptive data including authors, publication date, and status, as well as other information to characterize the study, including study design, country, type of intervention and outcome, population, and context[3]
- Methodological information, analysis method, and type of comparison (if relevant)
- Quantitative data for outcome measures, including outcome descriptive information, sample size in each intervention group, outcome means and standard deviations, and test statistics (e.g. t-test, F-test, p-values, 95% confidence intervals)
- Information on intervention design, including how the intervention incorporates participation, participant adherence, contextual factors, and programme mechanisms, including implementation fidelity

b. Critical appraisal

We will apply a critical appraisal tool to assess the trustworthiness of the impact evaluations included in the systematic review. Trustworthiness refers to the confidence that findings reported in the included impact evaluations were rigorous and credible and are likely to reflect the results of the evaluated interventions rather than the influence of the applied study design and research conduct. To assess the risk of bias in primary studies, we will adapt the Cochrane risk of bias tool for randomized and non-randomized studies (Sterne and others, 2016). This risk of bias tool has previously been used and adapted in international development reviews (Stewart and others, 2015; Langer and others, 2017). Sterne and colleagues (2016) used a domain-based risk of bias tool covering the following six indications of trustworthiness: (i) selection bias, (ii) confounding bias, (iii) bias due to departures from applied interventions, (iv) bias due to missing data, (v) bias due to measurement of outcomes, and (vi) bias due to selection of the reported result. Each bias domain will receive a low, moderate, high, or critical risk of bias rating, allowing for a transparent calculation of the overall risk of bias score for each study. Studies with a critical risk of bias will be excluded from the synthesis.

The critical appraisal tool used to assess studies for the systematic review is presented in Appendix 3. It will be piloted using a similar approach to that used to pilot the data extraction tool. Two reviewers will independently assess each study and then collaborate on a comparative review. A third reviewer will be consulted if these reviewers disagree about the risk of bias rating for a particular study.

[3] This information was already extracted in the development of EGM.

c. Methods for handling dependant effect sizes

i. Criteria for the determination of independent findings

Complex data structures are a common occurrence in meta-analyses of impact evaluations. There are numerous scenarios through which these complex structures can affect the meta-analysis. For example, several publications could originate from one study, or several studies could originate from the same data set. Some studies might have multiple treatment arms compared to a single control group. Other studies may report outcome measurements from several time points or use multiple outcome measures to assess related outcome constructs. Such cases yield statistically dependent effect size estimates (Borenstein and others, 2009).

The research team will assess the extent to which relationships exist across the studies included in the review and avoid double counting identical evidence by linking papers before data analysis. When several publications report on the same effect, effect sizes from the most recent publication will be used. The information provided in studies to support these assessments, such as sample sizes, programme characteristics and key implementing and/or funding partners, will be utilized.

We will extract effects reported across different outcomes or subgroups within a study. Where information is collected on the same programme for different outcomes at the same or different periods, information on the full range of outcomes over time will be extracted. Where studies report effects from multiple model specifications, we will adopt the author's preferred model specification. If this is not stated or is unclear, the specification with the most controls will be used. Where studies report multiple outcomes or evidence according to subgroups of participants, we will record and report data on relevant subgroups separately. Further information on the criteria for determining independent effect sizes is presented below.

We will deal with dependent effect sizes through data processing and selection techniques that utilize several criteria to select one effect estimate per study. When we have several publications reporting on the same study, we will use effect sizes from the most recent publication. For studies with outcome measures at different time points, we will follow De La Rue and others (2014) and synthesize outcomes measured immediately after the intervention (defined as one to six months) and at follow-up (longer than six months) separately. If multiple time points exist within these periods, we will adopt the most recent measure.

We anticipate that many of the interventions we address in our review will be ongoing programmes. We expect the follow-up will reflect a programme's duration rather than the time since the intervention. Where such studies report outcome measures at different time points, we will identify the most common follow-up period and include the follow-up measures that match this most closely in the meta-analysis. When studies include multiple outcome measures to assess related outcome constructs, we will follow Macdonald and others (2012) and select the outcome that reflects the construct of interest most accurately without referencing the results.

Studies may include multiple treatment arms with only one control group and the treatments representing separate treatment constructs. In such cases, we will calculate the effect size for treatment A versus control and treatment B versus control and will include them in separate meta-analyses according to the treatment construct. Where treatments A and B represent variations of the same treatment construct, we will apply the following approach. We will calculate the weighted mean and standard deviation for treatments A and B before calculating the effect size for the merged group versus the control group, following the procedures outlined in Borenstein and colleagues (2009, chapter 25). There may also be cases where different studies report on the same programme but use different samples (e.g. from different regions or separately for men and women). In such

instances, we will include both estimates, treating them as independent samples, provided effect sizes are measured relative to separate control or comparison groups.

ii. Effect size calculations

Using Excel, we will extract quantitative data for outcome measures, including outcome descriptive information, sample size in each intervention group, outcome means and standard deviations, and test statistics (e.g. t-test, F-test, p-values, 95% confidence intervals). Effect size data will be stored, and any necessary cleaning will be conducted in Excel. Following the screening and descriptive data extraction process, to ensure consistent coding, two reviewers will pilot the extraction tool for determining effect size. They will work independently on a random sample (10%) of included studies to test the tool across a range of the included impact evaluation designs and methods. We aim to achieve a minimum Kappa statistic score of 0.90 following a round of repeating the process for the tool to be finalized. After the piloting stage, individual reviewers will code the remaining studies and a third reviewer will check the extracted data.

An effect size expresses the magnitude (or strength) and direction of the relationship of interest (Valentine and others, 2015; Borenstein and others, 2009). We will extract data from each study to calculate standardized effect sizes for cross-study comparison wherever possible. For continuous outcomes comparing group means in a treatment and control group, we will calculate the standardized mean difference (SMD), or Cohen's *d*, its variance, and standard error (SE) using formulae provided in Borenstein and colleagues (2009). An SMD is a difference in means between the treatment and control groups divided by the pooled standard deviation of the outcome measure. Cohen's *d* can be biased in cases where sample sizes are small. Therefore, in all cases we will adjust *d* using Hedges' method, adjusting Cohen's *d* to Hedges' *g* using the following formula (Ellis, 2010):

$$g \cong d(1 - \frac{3}{4(n_1 + n_2) - 9})$$

Details of the appropriate formula for effect size calculations in reference to, and dependent on, the data provided in included studies are described in Appendix 4.

d. Data synthesis

Based on studies assessed to be sufficiently similar, we will combine studies using meta-analysis only when we identify two or more effect sizes using a similar outcome construct and where the comparison group state is judged as similar across the two (cf. the approach taken by Wilson and colleagues, 2011). We will combine studies in the same analysis when they evaluate the same intervention type and the same outcome type. Where there are too few studies or included studies are considered too heterogeneous in terms of interventions or outcomes, we will discuss the individual effect sizes along the causal chain. As programme theory of interventions suggests that there will be heterogeneity across studies, we will adopt inverse-variance weighted random effects meta-analytic models (Higgins and others, 2020) to account for this.

We will conduct separate analyses across the major outcome categories for each intervention type: knowledge, uptake and use outcomes, behavioural outcomes, development results, and impact outcomes that meet the inclusion criteria. Based on an analysis of the interventions we find, we attempt to further elaborate on the above pathway of change to the extent possible.

Whenever feasible, we aim to conduct moderator analyses to explain variations in effect sizes. Moderators are variables such as socioeconomic context and population characteristics, measured at baseline, that interact with treatment to change the outcome for each group (Pincus and others, 2011). Following the PROGRESS-PLUS approach (Oliver and others, 2017), we will use

moderators falling into three broad categories of extrinsic, methodological, and substantive characteristics. Specifically, these categories include:

- Extrinsic characteristics: funder of the study (e.g. non-governmental organization/civil society organization versus private sector versus government investments), publication type, publication date.
- Methodological characteristics: study design, risk of bias, evaluation period, length of intervention.
- Substantive characteristics: participant characteristics (gender, age, socioeconomic status), context (geographical setting), intervention type, intervention features, type of implementing agency.

We will use random effects meta-regression to investigate the association between moderator variables and heterogeneity of treatment effects (Borenstein and others, 2009) and subgroup analyses to investigate heterogeneity by treatment subgroups (e.g. men and women, poor and non-poor, and so on). If we do not have enough studies or data – we will discuss and explore the factors driving the heterogeneity of results narratively by conducting cross-case comparisons (Miles and Huberman, 1994).

e. Assessment of heterogenicity

To visibly examine variability in the effect size estimates, we will use forest plots to display the estimated effect sizes from each study along with their 95% confidence intervals. Subsequently, and acknowledging the limitations of quantification of heterogeneity and the different strengths of statistical approaches, we will conduct a heterogeneity test. Our test will consist of a calculation of the Q- statistic as a statistical test of heterogeneity (Hedges and Olkin, 1985) and a calculation of the i2 and Tau2 statistic to provide estimates of the magnitude of the variability across study findings caused by heterogeneity (Higgins and Thompson, 2002; Higgins and others, 2003; Borenstein and others, 2009).

f. Sensitivity analysis

To test the robustness of the results of the meta-analysis, we will conduct several sensitivity analyses. Broadly, this will involve collecting data on and assessing the sensitivity of findings to (i) the methods of the primary studies and (ii) the methods of the review. We anticipate the included studies will vary methodologically. Therefore, we will conduct sensitivity analyses to examine the influence of these variations on the summary measures to offer possible explanations for the differences between studies when interpreting the results. We will examine whether the results were sensitive to study design, the risk of bias associated with the study, the degree of missing/incomplete data, how outcomes are measured, and the timing of when they were measured. The main objective of the sensitivity analysis is to serve as a visual tool that allows informal comparisons to determine whether the results of our meta-analyses are sensitive to the methodological decisions of the review team. The sensitivity analyses will be carried out by adopting a one-way random effects analysis of variance (ANOVA) model calculated in EPPI-reviewer 4.

g. Strength of the evidence assessment

The last research step in the systematic review will be to conduct a Grading of Recommendations, Assessment, Development and Evaluations (GRADE) assessment to report on the overall strength of the evidence base and recommendations made based on the synthesis of the review. This step is distinct from the critical appraisal step. It considers additional factors to assess the overall body of the evidence and the reliability of the recommendations derived from it. Appendix 5 presents the GRADE tool with hypothetical decisions for illustration purposes.

D. CONCLUSION

There is currently a lack of rigorous empirical evidence on what could encourage or potentially bring about a change in human behaviour in a way that would be useful for reducing greenhouse gas emissions and adapting to the changed climate, especially in developing countries that are more vulnerable to the impacts of climate change. Against this backdrop, this review aims to assess and shed more light on how climate related actions can be implemented more effectively. The key objectives of this protocol for a systematic review are to identify, assess and synthesize evidence on the effectiveness of behavioural science interventions. These interventions include feedback, reminders, salience (communication), salience (design) and goal setting on environmental and development outcomes in developing countries. In fulfilling its objective, the review will facilitate the use of evidence in informing policy and practice decisions within the environmental and development fields, particularly climate mitigation and adaptation.

The overall evidence review uses a two-stage systematic approach. The first stage consists of an already completed EGM. The second stage consists of a systematic review and synthesis in compliance with the Campbell Collaboration's guidelines on producing EGMs and SRs. The systematic review and synthesis will be conducted on selected bodies of evidence contained in the EGM. An assessment will be undertaken of the effectiveness of the selected interventions in achieving the desired behavioural changes in developing countries' environmental and development outcome areas. Therefore, the systematic review will only include primary studies that measure the effects of interventions and whose design can reliably attribute observed effects to these applied interventions. Individual effects will be synthesized into overall estimates of treatment effects using statistical meta-analysis. This protocol outlines the data collection and analysis, including data extraction and management, critical appraisal of the evidence, methods for handling dependent effect sizes,[4] data synthesis,[5] heterogeneity assessment, sensitivity analysis and strength of evidence assessment.

E. DISTRIBUTION OF MAIN ROLES

Table 2. Distribution of main roles

PROJECT FUNDING, OVERSIGHT AND CO-CREATION	
Dr. Martin Prowse	Independent Evaluation Unit, Green Climate Fund
Deborah Sun Kim	Independent Evaluation Unit, Green Climate Fund
Yeonji Kim	Independent Evaluation Unit, Green Climate Fund
Elangtlhoko Mokgano	Independent Evaluation Unit, Green Climate Fund
Andreas Reumann	Independent Evaluation Unit, Green Climate Fund
Prof. Dr. Jyotsna Puri	International Fund for Agricultural Development
Dr. Romina Cavatassi	International Fund for Agricultural Development

[4] These include criteria for the determination of independent findings and calculation of effect sizes.
[5] These include meta-analysis and moderator analysis.

Project execution	
Africa Centre for Evidence	
Dr. Laurenz Langer	Project oversight and management Co-PI: Project oversight and management, stakeholder and client engagement, finance and reporting, drafting and finalization of outputs and deliverables. Research: Mapping and Synthesis Synthesis method lead: design of all research activities, tools development, and research staff training and support where relevant. Lead on formulating inclusion criteria, EGM framework development, meta-analysis, GRADE assessment, and Qualitative Comparative Analysis, if conducted
Prof. Ruth Stewart	Synthesis adviser: Ad hoc advice and technical inputs related to complex evidence synthesis, climate change and behavioural science issues encountered
Promise Nduku	Research Lead Research: Mapping and Synthesis Synthesis specialist: design and conduct of search strategy, screening, data extraction, and critical appraisal for both the EGM and the SR, lead on EGM visualization and effect size calculation
Tafadzwa Mutanha	Research: Mapping and Synthesis Research assistant: research support in accessing full-text articles, cataloguing data, collating background information, and editorial support
Sefora Rangoanana	Research: Mapping and Synthesis Research assistant: research support in accessing full-text articles, cataloguing data, collating background information, and editorial support
Content experts	
Dr. Benjamin Curtis	Subject input related to behavioural science on the following areas: scope of the EGM and SR including theory of change; inclusion criteria for the EGM and SR; framework development for the EGM; reviewing search strategy and output; data extraction variables for SR; interpreting synthesis results; output review including approach paper, EGM report and SR protocol and technical report; and stakeholder and client engagement
Dr. Caitlin Blaser Mapitsa	Content adviser: Ad hoc advice and technical inputs related to complex evidence synthesis, theory of change development, climate change, and behavioural science issues encountered
Jamie Robertsen	Subject input related to climate change on the following areas: scope of the EGM and SR including theory of change; inclusion criteria for the EGM and SR; framework development for the EGM; reviewing search strategy and output; data extraction variables for SR; interpreting synthesis results; output review including approach paper, EGM report and SR protocol
Samantha Booth	Content advisor: Ad hoc advice and technical inputs related to complex evidence synthesis, theory of change development, climate change, and behavioural science issues encountered

REFERENCES

Ashraf, Nava, Dean Karlan, Wesley Yin (2010). Commit to Saving: Don't Stop Until You Save Enough. Available at *https://www.bhub.org/project/commit-to-saving-dont-stop-until-you-save-enough/*.

Balmford, A., and others (2021). Making more effective use of human behavioural science in conservation interventions. *Biological Conservation*, vol. 261, No. 109256.

Behavioural Insights Team (2016). *2015-2016 Update Report.* Available at *https://www.bi.team/publications/the-behavioural-insights-teams-update-report-2015-16/*.

Binet, S., and others (2021). *Independent evaluation of the adaptation portfolio and approach of the Green Climate Fund*. Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund.

Bollinger, Bryan, Kenneth. T. Gillingham, and Marten Ovaere (2020). Field experimental evidence shows that self-interest attracts more sunlight. *Proceedings of the National Academy of Sciences*, vol. 117, issue 34, pp. 20503-20510.

Borenstein, M., and others (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex: John Wiley and Sons.

Byerly, Hilary, and others (2018). Nudging pro-environmental behaviour: evidence and opportunities. *Frontiers in Ecology and the Environment*, vol. 16, issue 3, pp. 159-168.

Carmody, Dennis, and Michael Lewis (2006). Brain activation when hearing one's own and others' names. *Brain Research*, vol. 1116, No. 1, pp. 153-158. Available at *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1647299/*.

Centre for Behaviour and the Environment (2020). Leveraging Social Norms to Reduce Household Energy Use. Case study.

Chapman, Gretchen, and others (2015). Goals and social comparisons promote walking behaviour. *Medical Decision Making*, vol. 36, issue 4, pp. 472-478. Available at *https://www.researchgate.net/publication/279727587_Goals_and_Social_Comparisons_Promote_Walking_Behavior*.

Cinner, Joshua (2018). How behavioral science can help conservation. *Science*, vol. 362, issue 6417, pp. 889– 890.

Datta, Saugato, and Sendhil Mullainathan (2014). Behavioral design: A new approach to development policy. *Review of Income and Wealth*, vol. 60, issue 1, pp. 7–35. Available at *https://onlinelibrary.wiley.com/doi/full/10.1111/roiw.12093*.

De La Rue, Lisa, and others (2014). School-based interventions to reduce dating and sexual violence: A systematic review. *Campbell Systematic Reviews*, vol. 10, issue 1, pp. 1-110. Available at *https://onlinelibrary.wiley.com/doi/full/10.4073/csr.2014.7*.

Duflo, Esther, Michael Kremer, and Jonathan Robinson (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review*, vol. 101, No. 6, pp. 2350–2390.

Flanagan, Ann Elizabeth, and Jeffery Clark Tanner (2016). Evaluating behavior change in international development operations: A new framework. Working Papers No. 2 (January). Washington, D.C.: World Bank. Available at *https://openknowledge.worldbank.org/handle/10986/25872*.

Gifford, Robert, Christine Kormos and Amanda McIntyre (2011). Behavioral dimensions of climate change: Drivers, responses, barriers, and interventions. *Wiley Interdisciplinary Reviews: Climate Change*, vol. 2, issue 6, pp. 801–827.

Global Commission on Adaptation (2019). Adapt now: A global call for leadership on climate resilience. Global Commission on Adaptation. Flagship reports (September). Washington, D.C.: Global Center on Adaptation and World Resources Institute. Available at *https://gca.org/reports/adapt-now-a-global-call-for-leadership-on-climate-resilience/*.

Hallsworth, Michael, and others (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, vol. 148, issue C, pp. 14–31.

Hasvold, Per E, and Richard Wootton (2011). Use of telephone and SMS reminders to improve attendance at hospital appointments: a systematic review. *Journal of telemedicine and telecare*, vol. 17, issue 7, pp. 358-364. Available at *https://journals.sagepub.com/doi/full/10.1258/jtt.2011.110707*.

Haynes, Laura, and others (2012). Test, learn adapt: Developing public policy with randomised controlled trials. *Behavioural Insights Team, Cabinet Office.* Available at *https://www.researchgate.net/publication/256031307_Test_Learn_Adapt_Developing_Public_Policy_with_Randomised_Controlled_Trials*.

Hedges, L.V. and Ingram Olkin (1985). Statistical methods for meta-analysis. *Journal of Educational Statistics*, vol. 20, No. 1. Available at *https://www.researchgate.net/publication/216811655_Statistical_Methods_in_Meta-Analysis*.

Higgins, Julian P.T., and others (2003). Measuring inconsistency in meta-analyses. *BMJ*, vol. 327, No. 7414, pp. 557-60.

Higgins, Julian P.T., and Simon G. Thompson (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, vol. 21, issue 11, pp. 1539-1558 (June).

Hollands, Gareth J, and others (2017). The TIPPME intervention typology for changing environments to change behaviour. *Nature Human Behaviour*, vol. 1, No. 8, pp. 1–9 (July). Available at *https://www.researchgate.net/publication/317559049_The_TIPPME_intervention_typology_for_changing_environments_to_change_behaviour*.

Intergovernmental Panel on Climate Change (2014). Summary for policymakers. In *Climate change 2014: Mitigation of climate change. Contribution of working group III to the fifth Assessment report of the Intergovernmental Panel on Climate Change*, Edenhofer, O., R. Pichs-Madruga, Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx, eds. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Available at *https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_summary-for-policymakers.pdf*.

Karlan, Dean, and others (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, vol. 62, No. 12, pp. 3,393-3,411.

Keef, Stephen P., and Leigh A. Roberts (2004). The meta-analysis of partial effect sizes. *British Journal Mathematical and Statistical Psychology*, vol. 57, issue 1, pp. 97-129 (December).

Keith, Wade, and Marcus Jennings (2015). Climate change and the global economy: Regional effects. Schroders. Available at *https://www.schroders.com/en/ch/asset-management/insights/economics/climate-change-and-the-global-economy-regional-effects/*.

Kremer, Michael, and others (2014). Source dispensers and home delivery of chlorine in Kenya. *Innovations for Poverty Action.* Available at*: https://www.poverty-action.org/study/source-dispensers-and-home-delivery-chlorine-kenya*.

Langer, Laurenz, and others (2018). Women in wage labour: A systematic review of the Effectiveness and Design Features of Interventions Supporting Women's Participation in Wage Labour in Higher-growth and/or Male-dominated Sectors in Low-and Middle-income Countries. Technical report. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. Available at *https://www.researchgate.net/publication/331993360_Women_in_wage_labour_A_systematic_review_of_the_Effectiveness_and_Design_Features_of_Interventions_Supporting_Women's_Participation_in_Wage_Labour_in_Higher-growth_andor_Male-dominated_Sectors_in_Low-*.

Lawlor, Blaine K., and Martin J. Hornyak (2012). Smart Goals: How the application of smart goals can contribute to achievement of student learning outcomes. *Developments in Business Simulation and Experiential Learning*, vol. 39, pp.259-267. Available at *https://journals.tdl.org/absel/index.php/absel/article/view/90*.

Locke, Edwin A. (1996). Motivation through conscious goal setting. Applied and Preventive Psychology, vol. 5, pp. 117-124. Available at *https://sci-hub.hkvisa.net/10.1016/s0962-1849(96)80005-9*.

Locke, Edwin A., and Gary Latham (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey." *American psychologist*, vol. 57, No. 9, pp. 705-717.

Macdonald, Geraldine, and others (2012). Cognitive-behavioural interventions for children who have been sexually abused: A systematic review. *Campbell Systematic Reviews*, vol. 8, issue 1, pp. 1-111. Available at *https://www.researchgate.net/publication/353619940_Cognitive-Behavioural_Interventions_for_Children_Who_Have_Been_Sexually_Abused_A_Systematic_Review*.

Miles, Matthew B., and A. Mechael Huberman (1994). *Qualitative data analysis*, 2nd ed. An expanded sourcebook. California, U.S.: SAGE. Available at *https://vivauniversity.files.wordpress.com/2013/11/milesandhuberman1994.pdf*.

Nisa, Claudia F., and others (2019). Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. *Nature Communications*, vol. 10, No. 4545, pp. 1–13. Available at *https://www.nature.com/articles/s41467-019-12457-2*.

Organisation for Economic Co-operation and Development (2017). Behavioural insights and public policy: Lessons from around the world. Paris: OECD Publishing. Available at *https://www.oecd.org/gov/regulatory-policy/behavioural-insights-and-public-policy-9789264270480-en.htm*.

Osbaldiston, Richard, and John Paul Schott (2012). Environmental sustainability and behavioral science: Meta-analysis of pro-environmental behavior experiments. *Environment and Behavior*, vol. 44, issue 2, pp. 257–299.

Pincus, Tamar, and others (2011). Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. *BMC medical research methodology*, vol. 11, article 14, Available at *https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-14*.

Polanin, Joshua R., and others (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, vol. 86, issue 1, pp. 207-236.

Sánchez-Meca, Julio, Fulgencio Marín-Martínez, and Salvador Chacón-Moscoso (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological methods*, vol. 8, No. 4, pp. 448-467.

Schot, Johan, Laur Kanger, and Geertb Verbong (2016). The roles of users in shaping transitions to new energy systems. *Nature Energy*, vol. 1, No. 5, pp. 1–7. Available at *https://www.researchgate.net/publication/302065665_The_roles_of_users_in_shaping_transitions_to_new_energy_systems*.

Snilstveit, Birte, and others (2019). Incentives for climate mitigation in the land use sector - The effects of payment for environmental services on environmental and socioeconomic outcomes in low-and middle-income countries: A mixed-methods systematic review. *Campbell Systematic Reviews*, vol. 15, issue 3. Available at *https://onlinelibrary.wiley.com/doi/epdf/10.1002/cl2.1045*.

Solomon, Susan, and others, eds (2007). *Climate change 2007: The physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Available at *https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-frontmatter-1.pdf*.

Stern, Paul C. (2020). A reexamination on how behavioral interventions can promote household action to limit climate change. *Nature communications*, vol. 11, article 918, pp. 1-3.

Stern, Paul C., and others (2016). Opportunities and insights for reducing fossil fuel consumption by households and organizations. *Nature Energy*, vol. 1, No. 5, article No. 16043, pp. 1–6.

Sterne, Ac Jonathan (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, vol. 355.

Stoknes, Per Espen (2014). Rethinking climate communications and the "psychological climate paradox". *Energy Research and Social Science*, vol. 1, pp. 161–170.

Szaszi, Barnabas, and others (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, vol. 31, issue 3, pp. 355–366. Available at *https://onlinelibrary.wiley.com/doi/10.1002/bdm.2035*.

Thaler, Richard H., and Cass R. Sunstein (2009). "Nudge: Improving decisions about health, wealth, and happiness". New York, NY, USA: Penguin Randon House.

United Nations Environment Programme (2017). The emissions gap report 2017. Synthesis report. Nairobi. Available at *https://wedocs.unep.org/bitstream/handle/20.500.11822/22070/EGR_2017.pdf*.

Valentine, Jeffrey, Ariel Aloe, and Timothy Lau (2015). Life after NHST: How to describe your data without "p-ing" everywhere. *Basic and Applied Social Psychology*, vol. 37, No. 5, pp. 260-273.

Williamson, Katie, and others (2018). Climate change needs behavior change: Making the case for behavioral solutions to reduce global warming. Center for Behavior and the Environment (rare). Arlington, VA. Available at *https://rare.org/wp-content/uploads/2019/02/2018-CCNBC-Report.pdf*.

Wilson, David B., and others (2011). Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. *Campbell Systematic Reviews*, vol. 7, issue 1, pp. 1-53. Available at *https://www.researchgate.net/publication/265409241_Use_of_DNA_Testing_in_Police_Investigative_Work_for_Increasing_Suspect_Identification_Arrest_Conviction_and_Case_Clearance*.

Zhao, Ya Yun, and others (2019). The effect of text message reminders on medication adherence among patients with coronary heart disease: A systematic review and meta-analysis. *Medicine*, vol. 98, issue 52 (December). Available at *https://journals.lww.com/md-journal/Fulltext/2019/12270/The_effect_of_text_message_reminders_on_medication.8.aspx*.

APPENDICES

Appendix 1. DESCRIPTION OF INTERVENTION TYPES

The systematic review informed by this protocol will determine if feedback, reminders, salience (communication), salience (experience design), and goal setting interventions achieve their desired environmental and development outcomes in developing countries. Detailed descriptions of these intervention types are as follows:

Reminders

Reminders call people's attention to something and encourage them to take certain actions. Reminders typically involve sending a message at a timely moment with a specific call to action. For example, emailing or texting a reminder to someone 24 hours ahead of their medical appointment. Reminders help counteract one of the cognitive limitations human beings face – the fact that we have limited attention and memory. Even when completing a behaviour or task is important (like filing taxes or going to the doctor), it may be forgotten because we have other obligations that compete for our attention. Reminders anticipate that people are likely to forget and help direct their attention to an action that needs to be taken, making it more salient and bringing it to the mind's forefront at the right moment.

Reminders can be delivered as 'pure' reminders that remind a person to do something at the appropriate moment. They can also be coupled with other behavioural science interventions, such as using gain or loss framing or including social norms in the messaging. They can be delivered as one-off reminders for once only behaviours, like attending an appointment. Or they can be repeated reminders for repeated behaviours, like taking medication. Reminders have been used to successfully influence a variety of behaviours, such as increasing savings (Karlan and others, 2016), attending appointments (Hasvold and Wootton, 2011), and adhering to medical treatments (Zhao and others, 2019). For example, when the Court Service of the United Kingdom sent text messages reminding people of the closing date for the payment of their outstanding fines, fine payment amounts doubled. Also, personalizing the reminder message increased fine payment amounts by a further 45 per cent (Service and others, 2014).

Feedback

Feedback interventions provide information, often tracked over time, about a particular behaviour. Feedback generally indicates how 'well' someone is doing in relation to a target, their own past performance, or others' behaviour. It may also outline the consequences of the behavioural trajectory.

Feedback interventions are effective at shifting behaviour because they draw attention to the behaviour and put it into context by providing a benchmark. For example, by establishing a benchmark then tracking progress towards achieving it, feedback interventions can encourage continued progress as well as the behaviours that drove that behaviour. Feedback interventions can also help people understand the consequences of their behaviours. For example, by tracking direct results of actions.

Optimal feedback is real-time or immediate, and most effective for people who are underperforming – such as people who use too much electricity. Feedback interventions can, however, backfire for those already performing well in relation to others. For example, someone who learns they are using less electricity than their neighbours may increase their electricity consumption). Feedback interventions typically require tracking a behaviour over time and are best suited to influencing repeated behaviours.

Feedback has been used to reduce speeding (ACT Government, 2020), decrease energy use, and increase recycling. (Center for Behaviour and the Environment, 2020). An energy company, OPower, has used personalized feedback in household energy reports to help customers reduce energy consumption. Their energy reports provide a simple bar graph showing a household's energy consumption in comparison to neighbours', including 'energy efficient' neighbours. This intervention reduced energy bills by an average of 1.5 – 2.5 per cent in the first two years of its implementation (Center for Behaviour and the Environment, 2020).

Salience (communication)

Salience (communication) interventions improve the ease and accessibility of adopting behaviours by making information or choices more prominent and relevant when communicated to people. These interventions typically focus on the messaging's content. They are distinct from reminders which focus on timely delivery. Simply sending out a communication (e.g. sending a letter or email) does not classify as a salience (communication) intervention. The communication needs to be tailored in a way that increases its salience. This could include things like making the content more relevant to an individual (e.g. personalizing it by using an individual's name), making it clear that the communication is important (e.g. using a big red stamp on a letter), making it easier for people to understand what needs to be done (e.g. laying out specific steps), or making it easier for people to carry out the action (e.g. including the phone number someone needs to call).

Salience (communication) interventions are effective because they increase the likelihood that people will pay attention (Carmody and Lewis, 2006) and understand what they are being asked to do. This can be applied to any form of communication. For example, a trial in the UK found that adding a person's name to a text message for collecting overdue fines increased the number of people making payments by 10 percentage points in comparison to a standard letter, and by 27.8 percentage points over those who received no text (Haynes and others, 2012). A similar trial used a red 'Pay Now' stamp on notices about fines which led to a 3.1 percentage point increase in payment rates (Behavioural Insights Team, 2016).

Salience (experience design)

Interventions classified under salience (experience design) target how individuals interact with their physical and/or digital environment. They typically involve changing aspects of a process, such as arranging facilities or options so that they are either more prominent, accessible, and easy to prompt a particular behaviour, or less prominent, accessible, or easy, to discourage a particular behaviour. They are distinct from salience (communication) interventions as they focus on how people interact with their environment and not on messaging content.

Salience (experience design) interventions are effective because they remove or add frictions to carrying out behaviours. People are extremely sensitive to frictions. Small, seemingly minor details that make a task more effortful have a disproportionately large effect on whether people complete a task. Salience (experience design) interventions leverage this tendency to make it more or less likely that someone takes an action.

Salience (experience design) can take a variety of forms. Examples include changing the ordering of items on menus (people tend to choose the first and last options more frequently), placing healthy food first in cafeteria lines, simplifying forms to make it more likely that people complete them, or reducing the number of steps in a process. In Kenya, a trial was run to test whether installing chlorine dispensers directly at water sources could increase the use of chlorine in treating drinking water. This simple intervention increased chlorine usage by 53 percentage points (Kremer and others, 2014).

Goal setting

Goal setting interventions help individuals consider what their priorities are, then specify a series of goals they would like to achieve. The goals need to be specific and are typically specified by the individual or group whose behaviour is being influenced but may also be externally determined. For example, in health applications, individuals might set their own targets for weight loss, or they might be given a set of medically validated 'best practice' targets. These interventions are often coupled with a planning process and may also be combined with other behavioural insights (such as mental contrasting, implementation intentions, endowed progress, commitment devices, or feedback) to encourage achievement of the goal.

Goal setting interventions are effective because they direct attention towards goal-relevant activities and motivate goal-oriented behaviours (Locke and Latham, 2002). Self-set goals, particularly those that are more difficult, are more likely to lead to commitment and action (Locke, 1996). Goal setting interventions have been used to successfully improve student learning outcomes (Lawlor and Hornyak, 2012), increase savings (Ashraf and others, 2010), and increase exercise (Chapman and others, 2015). An example of the latter occurred at a university in North America where physical activity among staff members increased when they were given daily step (walking) goals. Those receiving a high goal walked on average 1,912 more steps per day than those given a low goal (Chapman and others, 2015). An example from the environmental field is setting a goal for recycling or using low-emission transport to contribute to climate change mitigation (Nisa and others, 2019).

Appendix 2. DATA EXTRACTION TOOL

DESCRIPTION	QUESTION
Date when form was completed	*Date when form was completed*
ID of person extracting data	*ID of person extracting data*
Report Identification	
Publication title	*Title of publication*
Publication ID	*EPPI ID*
Author details	*Surname of first author*
Publication date	*Year (letter - if more than one study from that author and that year)*
Publication type	*What is the impact evaluation publication type?* ☐ *Academic journal article* ☐ *Research report* ☐ *Government report* ☐ *Dissertation / thesis* ☐ *Online book chapter*
Funding agency name	*Who is funding the evaluation/study? Please add name of the agency funding the evaluation.*
Funding agency type	*Type of agency funding the evaluation/study:* ☐ *Academic institution* ☐ *Charitable or private foundation* ☐ *For-profit firm* ☐ *Government agency* ☐ *International aid agency* ☐ *International financial institution* ☐ *Non-profit organization* ☐ *Not specified*
Independence of evaluation	*What level of independence is there between the implementing agency and study team?* ☐ *Funding and author team independent of implementers/ funders of programme* ☐ *Funding independent of implementers/ funders of programme, but includes authors from funder/ implementer* ☐ *Evaluation funded and undertaken by funders/ implementers* ☐ *Unclear*
Independent data collection	*Has the data been collected by an independent party?* ☐ *Yes* ☐ *No* ☐ *Not clear*
Conflict of interest	*Is there a potential conflict of interest associated with the study which could influence the collected/reported results? (e.g. Is there a declaration of conflict*

DESCRIPTION	QUESTION
	of interest? Is any of the authors related in any way to the funding or implementing institution?) ☐ *Yes* ☐ *No* ☐ *Not clear*
Comments on conflict of interest	*Please add reason for your answer to whether there is a conflict of interest.*
Language of publication	*Language of publication of the impact evaluation (e.g. Spanish, English etc)*
Other methods	*If the impact evaluation addresses questions other than effectiveness, note the questions and methods used here.*
Linked studies	*If there is any study linked to this one, add reference.*
Context	
Country	*List countries the study was conducted in.*
Detailed location	*If provided, give detailed information on where the study took place within a country (e.g. regions/districts covered).*
World Bank Region	*Select region(s) the study was conducted according to the World Bank. For more info on region classification see http://data.worldbank.org/country.*
World Bank Income category	*Select the World Bank income classification of the country at the time of the study.*
Sector	*Choose sector options below:* ☐ *Agriculture* ☐ *Education* ☐ *Energy and extractives* ☐ *Forestry* ☐ *Financial* ☐ *Industry and Trade/Services* ☐ *Information and Communication* ☐ *Public Administration* ☐ *Transportation* ☐ *Water sanitation and hygiene (WASH)* ☐ *Environmental and disaster management*
Intervention information	
Programme or project name	*State the programme or project name. If no name, then list the location.*
Study design	*Select the type of study:* ☐ *Randomized Control Trial (RCT)* ☐ *Regression discontinuity* ☐ *Matching/ Propensity Score Matching (PSM)* ☐ *Instrumental Variable/2SLS* ☐ *Difference in Difference* ☐ *Interrupted Time series analysis* ☐ *Controlled Before and After* ☐ *Heckman*

DESCRIPTION	QUESTION
	☐ *Fixed effects or random effects estimation* ☐ *Natural experiment*
Estimation methods	*Brief description of the estimation methods.*
Commentary on methods (if multiple methods are selected)	*State here if multiple methods are selected.*
Multicomponent intervention	*Is more than one intervention subcode applied to this intervention?* *- If yes, go to question 2.* *- If no, code as "No".* *Is each intervention subcode evaluated independently (i.e. separate effect sizes estimated for each intervention subcomponent, e.g. 2x2 design, separate evaluations reported in 1 study)?* *- If yes, code as "Multiple components, but evaluated separately".* *- If no, code as "Multiple components, not evaluated separately".* ☐ *Multiple components, but evaluated separately* ☐ *Multiple components, not evaluated separately* ☐ *No*
Number of treatment arms	*State the number of treatment arms.*
Treatment ID	*Please create an ID for each treatment of the intervention. 'Treatment' is defined here by 'treatment arms' (i.e. the combination of intervention components received by an arm of the evaluation).* *For example, in a case where there are three (3) intervention components A, B, and C, with two (2) treatment arms A+B and A+C, this would be coded on separate rows as:* *Treatment 1 Component A* *Treatment 1 Component B* *Treatment 2 Component A* *Treatment 2 Component C* *In cases where the intervention is the same (e.g. A+B and A+B), but the delivery mechanism is different (e.g. by community elders vs. by teachers), code as separate treatments.* *WHEN a study does not have a 'pure control', in which the comparison arm receives some intervention component, that comparison is coded as another treatment arm, even if there are no outcomes measured by that arm as a treatment.*
Component ID	*Please create a component ID for each component of the intervention.* *Component IDs need to be consistent across treatments. For example, if a component is repeated across treatments, it should have the same component ID.* *For example, in a case where there are three (3) intervention components A, B, and C, with two (2) treatment arms A+B and A+C, this would be coded on separate rows as:* *Treatment 1 Component A* *Treatment 1 Component B* *Treatment 2 Component A* *Treatment 2 Component C* *Component IDs should be captured alphabetically.*

DESCRIPTION	QUESTION
Intervention type[6]	*Select the intervention type:* *How* ☐ *Checklists* ☐ *Reduce hassles* ☐ *Rules of thumb* ☐ *Commitment devices* *Why* ☐ *Micro incentives* ☐ *Group incentives* ☐ *Lotteries* ☐ *Anchoring* ☐ *Framing devices* *Who* ☐ *Identity priming* ☐ *Public commitments* ☐ *Social norms* ☐ *Social benchmarking* ☐ *Cognitive behavioural therapy* *When* ☐ *Reminders** ☐ *Planning prompts* ☐ *Feedback* *Which* ☐ *Active choice* ☐ *Salience (communication)* ☐ *Salience (experience design)* ☐ *Goal setting* ☐ *Defaults*
Other (add new if does not fit existing categories)	*If you are certain the intervention does not fit within any of the previously defined classifications of behavioural science interventions, code the intervention here, otherwise leave blank. When developing a name, either use description from the study or if unclear code it as a non-behavioural science intervention.*
Description of Intervention(s)	*Write a short paragraph to describe the intervention type and characteristics. The description should be as detailed as possible. Add page numbers.*
Objectives of intervention	*State any objectives stated in study or other document.*
Scale of implementation	*At which level what the intervention implemented?* ☐ *Individual* ☐ *Household* ☐ *Firm*

[6] The review focuses on feedback, reminders, salience(communication), salience (experience design) and goal setting interventions which can be delivered as single interventions or in combination with other behavioural science interventions.

DESCRIPTION	QUESTION
	☐ *Community* ☐ *District/region*
What intervention (if any) did the comparison group receive?	☐ *No treatment* ☐ *As usual* ☐ *Alternative Intervention* ☐ *Other* ☐ *Unclear*
Intervention implementing agency name	*Who is implementing the intervention? State the name (and department) of the implementing agency*
Intervention implementing agency type	*Type of agency for the implementation of the intervention:* ☐ *Academic institution* ☐ *Charitable or private foundation* ☐ *For-profit firm* ☐ *Government agency* ☐ *International aid agency* ☐ *International financial institution* ☐ *Non-profit organization* ☐ *Not specified*
Intervention funding agency name	*Who is funding the intervention? State the name (and department) of the funding agency.*
Intervention funding agency type	*Type of funding/financial institution for the implementation of the intervention* ☐ *Academic institution* ☐ *Charitable or private foundation* ☐ *For-profit firm* ☐ *Government agency* ☐ *International aid agency* ☐ *International financial institution* ☐ *Non-profit organization* ☐ *Not specified*
Intervention target group	*What were the characteristics of the beneficiaries targeted by the intervention? were the characteristics of beneficiaries used to target the intervention? Open answer.*
Target population gender	*Indicate the gender of the targeted population:* ☐ *Female* ☐ *Male* ☐ *Female and male* ☐ *Unclear*
Target population age	*Indicate the population either* ☐ *Children <18* ☐ *Young adults (18-35)* ☐ *Adults (36-65)* ☐ *Elderly (65+)*

DESCRIPTION	QUESTION
	☐ *Mixed* ☐ *Not specified*
Target population income	*Indicate the target population income* ☐ *Low* ☐ *Middle* ☐ *Diverse* ☐ *Not specified*
Target population living environment	*State the target population living environment between* ☐ *Rural* ☐ *Urban* ☐ *Both*
Targeting methods	*How were beneficiaries targeted for the programme (e.g. how was the targeting implemented)?*
Target population specific restrictions	*Please provide details. Please provide details. In some instances, the target population is restricted to exclude population members that are difficult or impossible to interview.*
Intervention start	*Start date (if not stated, state study date) of intervention.*
Intervention end	*State end date (if ongoing state ongoing).*
Intervention length /exposure to intervention (in months)	*Start intervention length (months).*
Evaluation period (in months)	*The total number of months elapsed between the end of the intervention and the point at which an outcome measure is measured post-intervention, or as a follow-up measurement. If less than one month, use decimals (e.g. one week would be.25)*
Consideration of equity	*Does the study consider equity?* ☐ *Yes* ☐ *No*
Equity focus[7]	*How does the study consider equity?* ☐ *Intervention targets vulnerable population* ☐ *Subgroup analysis by sex* ☐ *Subgroup analysis (other than sex)* ☐ *Heterogeneity analysis (other than subgroup)* ☐ *Equity-sensitive analytical framework* ☐ *Equity-sensitive methodology* ☐ *Equity-sensitive research process* ☐ *Measures effects on an inequality outcome* ☐ *Research ethics informed by equity*
Equity dimension	*What dimension(s) of equity does the study consider?* ☐ *Age (e.g. old or young age but only if it provides arguments)* ☐ *Conflict-affected*

[7] The 3ie equity coding protocol and guidance is publicly available at *https://www.3ieimpact.org/sites/default/files/2021-11/DEP_Gender_Equity_Protocol-DEP.pdf*.

DESCRIPTION	QUESTION
	☐ *Culture (includes language)* ☐ *Disability (medical, physical, neurological, mental disorders)* ☐ *Education* ☐ *Ethnicity* ☐ *Head of household (female or male)* ☐ *HIV/AIDS (people with or at risk of HIV)* ☐ *Land size* ☐ *Land ownership* ☐ *Place of residence (rural, urban, peri-urban, informal dwellings)* ☐ *Refugees* ☐ *Religion* ☐ *Socioeconomic status (income or poverty status)* ☐ *Social capital* ☐ *Sex (includes the use of the term gender meaning the biological sex of a person)* ☐ *Sexual orientation* ☐ *Sexual identity* ☐ *Other (vulnerable groups not typified by any of the above). Answers might include orphans, sex workers, survivors of sexual violence etc.* ☐ *Not applicable*
Process and implementation	
Information about programme take-up	*Is there any information about programme take-up? Take-up refers to participation in a programme among those who are eligible.* *Commentary by authors should be used when information on programme take / up etc. is not backed up by some sort of research / when the authors do not report that/how they collected data to assess these areas.* ☐ *Yes, commentary from author* ☐ *Yes, formally assessed* ☐ *No*
Methods of assessing take-up	*Which methods are used to assess programme take-up?* ☐ *Observation by intervention staff* ☐ *Reporting by participants* ☐ *Other* ☐ *Commentary from author* ☐ *Not measured*
Results of the take-up assessment	*What is the result/information provided of the assessment of programme take-up?* *Open answer*
Information about programme adherence (among beneficiaries)	*Is there any information about programme adherence (how well the participants stuck to the programme requirements) among beneficiaries?* *Commentary by authors should be used when information on programme adherence etc. is not backed up by some sort of research or when the authors do not report that/how they collected data to assess these areas.* ☐ *Yes, commentary from author*

DESCRIPTION	QUESTION
	☐ *Yes, formally assessed* ☐ *No*
Methods of assessing adherence	*Which methods are used to assess programme adherence for beneficiaries? This includes dropout rates and adherence to appointments, etc.* ☐ *Observation by intervention staff* ☐ *Reporting by participants* ☐ *Other* ☐ *Commentary from author* ☐ *Not measured*
Results of the adherence assessment	*What is the result/ information provided of the assessment of programme adherence?* *Open answer*
Information about implementation fidelity/intervention delivery quality (among implementers)	*Is there any information on implementation fidelity/ intervention delivery quality?* *Commentary by authors should be used when information on programme adherence etc. is not backed up by some sort of research / when the authors do not report that/how they collected data to assess these areas.* ☐ *Yes, commentary from author* ☐ *Yes, formally assessed* ☐ *No*
Methods of assessing intervention fidelity	*Which methods are used to assess implementation fidelity/ intervention delivery quality by the implementing partner:* ☐ *Observation by intervention staff* ☐ *Reporting by participants* ☐ *Other* ☐ *Commentary from author* ☐ *Not measured*
Results of the intervention fidelity assessment	*What is the result/ information provided of the assessment of implementation fidelity/ intervention delivery quality?* *Open answer*
Incentives	*Were incentives provided to intervention participants?* ☐ *Yes* ☐ *No* ☐ *Not clear*
Other descriptions of process/implementation factors	*Any other description of process / implementation factors not covered above* *Open answer*
Results	*Report here any material relevant to causal mechanisms and barriers and enablers.* *Open answer*
Cost	*Are any unit cost data / cost-effectiveness estimates provided?* ☐ *Return on investment analysis* ☐ *Cost-effectiveness* ☐ *Cost benefit*

DESCRIPTION	QUESTION
	☐ *Cost only* ☐ *No cost data*
Cost details	*If yes, report any details of unit cost and/or total cost. Please also report the year and currency.*
External validity	
Length of study	*Length of study in months (Where study length is not reported, code as length of intervention and include a note in brackets)* *Number of months, if not reported N/A*
Efficacy or effectiveness trial	*Was the intervention implemented under "real world" conditions? By real world we mean a programme implemented independently of the evaluation, either by government, non-governmental organization, or international agency* ☐ *Yes* ☐ *No* ☐ *N/A*
Personnel implementing the programme	*Who was in charge of implementing the programme?* ☐ *PI/ researchers (study authors)* ☐ *Implementing agency staff* ☐ *External agency (e.g.: survey firm)* ☐ *Others* ☐ *Not clear*
Author discussion of external validity	*Do the authors discuss or explicitly address generalisability / applicability?* ☐ *Yes* ☐ *No*
Outcome information	
Outcome type[8]	*Select the outcome type:* *Knowledge, uptake and use outcomes* ☐ *Know of intervention* ☐ *Take part in intervention* ☐ *Acquire knowledge* ☐ *Change attitudes* *Behavioural outcomes* ☐ *Start behaviour* ☐ *Increase behaviour* ☐ *Decrease behaviour* ☐ *End behaviour* ☐ *No change in behaviour* *Development results* ☐ *Enhance Equity* ☐ *Support resource conservation* ☐ *Changing technologies*

[8] All selectable outcome options are outlined in the approach paper and the EGM report. See *https://ieu.greenclimate.fund/evidence-review/behavioural-science.*

DESCRIPTION	QUESTION
	☐ *Improve health Improve income and livelihoods* ☐ *Sustainable waste management* ☐ *Sustainable supply chain management and transport* *Impact* ☐ *Mitigation* ☐ *Adaptation*
Outcome indicator description	*Extract the exact name of the indicator being used as the dependent variable in the analysis. Use this open answer field to enter, in the author's own words, a description of the outcome, in a sentence or so. Be selective and concise with the excerpts being transcribed here as to ensure accurate and precise descriptions of the outcome. Include page numbers with every excerpt extracted.*
Outcome timing	☐ *Less than 1 year* ☐ *1 to 3 years* ☐ *More than 3 years* ☐ *Not clear*
Timing of outcome measurement	☐ *Only after* ☐ *Before and after* ☐ *Not clear*
Unintended outcomes	*State any unintended outcomes highlighted in the study.*
Effective size calculations	
Treatment ID	*Indicate the relevant treatment ID linked to the relevant effect size.*
Outcome type	*Select the outcome used to extract effect size data.*
Post-intervention or change from baseline?	☐ *Post-intervention* ☐ *Change from baseline*
Nature of the measures/estimate type	*Type of data for this effect size:* ☐ *Continuous* ☐ *Dichotomous outcome - proportions* ☐ *Hand calculated data* ☐ *Regression data*
Direction of the effect	☐ *Effect favours treatment* ☐ *Effect favours comparison* ☐ *Zero effect* ☐ *Unclear*
Reverse sign (i.e., decrease is good)	*Record no if an increase is good, record yes if a decrease is good and the sign needs to be reversed.* ☐ *Yes* ☐ *No*
Unit of analysis	☐ *Individual* ☐ *Household* ☐ *Firm* ☐ *Community*

DESCRIPTION	QUESTION
	☐ *District/region* ☐ *Unclear*
When measuring this outcome were there any differences between the treatment group participants and the comparison?	☐ *Yes* ☐ *No*
Effect is statistically significant?	☐ *Yes* ☐ *No* ☐ *Unclear*
Treatment sample size	*Insert treatment sample size here.*
Control sample size	*Insert control sample size here.*
Subgroup	*Is this analysis of a subgroup?* ☐ *Yes* ☐ *No*
If yes to subgroup, describe the subgroup if applicable	*Free text, describe the subgroup if applicable (e.g. boys, girls).*
Source	*Which page(s) contain the effect size data? Note the page number, table number, column, and row used to extract the data.*
The following group of questions applies only if Nature of the Measures is "Continuous"	
Treatment group mean	*Insert numerical value.*
Comparison group mean	*Insert numerical value.*
Are means reported above adjusted?	☐ *Yes* ☐ *No*
Treatment group standard deviation	*Insert numerical value.*
Comparison group standard deviation	*Insert numerical value.*
Treatment group standard error	*Insert numerical value.*
Comparison group standard error	*Insert numerical value.*
t-value from an independent t-test	*Insert numerical value.*
The following group of questions applies only if Nature of the Measures is "Dichotomous"	
Treatment group number of participants who experienced a change	*Insert numerical value.*
Comparison group number of participants who experienced a change	*Insert numerical value.*
Treatment group proportion of participants	*Insert numerical value.*

DESCRIPTION	QUESTION
who experienced a change	
Comparison group proportion of participants who experienced a change	*Insert numerical value.*
Are the proportions above adjusted for pre-test variables?	☐ *Yes* ☐ *No*
Logged odds ratio	*Insert numerical value.*
Standard error of logged odds ratio	*Insert numerical value.*
Logged odds ratio adjusted?	☐ *Yes* ☐ *No*
Chi-square with *df*=1 (2 by 2 contingency table)	*Insert numerical value.*
Correlation coefficient	*Insert numerical value.*
The following group of questions applies only if Nature of the Measures is "Hand Calculated Data"	
Hand calculated d-type effect size	*Insert numerical value.*
Hand calculated error of the d-type effect size	*Insert numerical value.*
Hand calculated odds ratio effect size	*Insert numerical value.*
Hand calculated odds ratio standard error	*Insert numerical value.*
Intermediate outcomes or themes (knowledge, skills)	*State intermediate outcomes or themes here.*
Questions applying to all studies	
Are there results coming from regressions?	☐ *Yes* ☐ *No*
Sample size	*Insert sample size here.*
The following group of questions applies only if there are results coming from regressions	
Method: Econometric model?	*State the econometric model*
Standard deviation effect	*Insert numerical value.*
Effect (mean)	*Insert numerical value.*
Controls	*Insert numerical value.*
Standard deviation: Y	*Insert numerical value.*
Standard deviation: X	*Insert numerical value.*
β (beta)	*Insert numerical value.*
Standard error β (beta)	*Insert numerical value.*
Degrees of freedom	*Insert numerical value.*

DESCRIPTION	QUESTION
Data type	☐ *Panel* ☐ *Cross-section* ☐ *Time series*

Appendix 3. CRITICAL APPRAISAL TOOL

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
(If randomized control trial, start after confounding bias. For all other study designs, start here.) **I. Bias in selection of participants into the study** *Are participants selected in a way that minimizes selection bias?*[9] **Appraisal indicators** Consider whether:						
1)	*There is an adequate description of how and why the sample was chosen (i.e., identified/selected/recruited).*					
2)	*There is adequate sample size to allow for representative and/or statically significant conclusions.*					
3)	*Participants in the control*[10] *group were sampled from the same population as that of the treatment.*					
4)	*The group allocation process minimized the potential risk of bias (e.g., using computer algorithms).*					
5)	*The selection of participants for the study (or the analysis) is based on participant characteristics observed after the start of the intervention.*					
Low risk of bias	Moderate risk of bias	High risk of bias	Critical risk of bias	Worth continuing? Y/N		
II. Bias due to confounding *Is confounding potentially controllable in the context of this study?* **Appraisal indicators** Consider whether:						

[9] Selection bias can occur both in the way that individuals are accepted for participation in a study and in the way that 'treatment' is assigned to individuals once they have been accepted into a study. This section deals with both these understandings of selection bias.
[10] The terms 'control' and 'comparison' group refer to any group with the treatment of interest is compared and is presumed to represent conditions in the absence of that treatment, whether it is true random or not.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
1) There is potential for confounding the effect of the intervention in this study. If yes, provide examples of confounding domains in the comment box.[11]						
2) Where matching was applied and, if so, whether it featured sufficient criteria.[12]						
3) Where relevant, the authors conducted an appropriate analysis that is controlled for all potential/remaining critical confounding domains after matching had been applied.						
4) The authors avoided adjusting for variables identified after the intervention has been administered.						
5) The treatment and control group are comparable after matching/controls have been completed. Select one of the following: ☐ *No statically significant differences* ☐ *Statistically significance difference* ☐ *Negligible descriptive differences* ☐ *Significant descriptive differences*						
Low risk of bias	Moderate risk of bias	High risk of bias	Critical risk of bias	Worth continuing? Y/N?		
(If randomized control trial, skip I + II (above) and start here.) **III. Bias due to confounding (because of ineffective randomization)** *Is allocation of treatment status truly random?* **Appraisal indicators** Consider whether:						
1) Eligibility criteria for study entry are specified.						
2) There is a clear description of randomization process and whether the methods are robust.						
3) The unit of randomization and number of participants is clearly stated (pay special attention to treatment and control locations/balance).						

[11] Confounding domains are those for which, in the context of this study, adjustment is expected to lead to an important change in the estimated effect of the intervention.
[12] Matching can be done on the calculated propensity score or covariates. If the latter, it should ideally be done on the pre-test measures and other characteristics, such as demographic. Answer 'no' if the study only matched on pre-test measures of some or all variables used later as outcome measures or matched only on end line characteristics.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
4) Characteristics of both baseline and end line sample are provided13 and at end line the treatment and control group are comparable. Select one of the following: ☐ *No statically significant differences* ☐ *Statistically significance difference* ☐ *Negligible descriptive differences* ☐ *Significant descriptive differences*						
Low risk of bias	Moderate risk of bias	High risk of bias	Critical risk of bias	If critical risk of bias, treat as non-random study		
IV. Bias due to departures from intended interventions *Was the intervention implemented as laid out in the study protocol?* **Appraisal indicators** Consider whether:						
1) *The critical co-interventions were balanced across intervention and control groups*						
2) *Treatment switches were low enough to not threaten the validity of the estimated effect of the intervention.*						
3) *Implementation failure was minor and unlikely to threaten the validity of the estimated effect of the intervention.*						
4) *It is possible that the intervention was taken by the controls (contamination and possible crossing-over).*[14]						
5) *It is possible that knowledge of group allocation affects how the two study groups are treated during delivery and evaluation of the intervention.*[15]						
Low risk of bias	Moderate risk of bias	High risk of bias	Critical risk of bias	Worth continuing? Y/N		
V. Bias due to missing/incomplete data (attrition) *Are the intervention and control groups free of critical differences in participants with missing/incomplete data?*						

[13] Preferable condition: An RCT with appropriate randomization procedure can be included without showing baseline data. As both experimental groups can be assumed to be equal as baseline by design.
[14] Whilst challenging in terms of estimating impact, spill overs might be an important finding.
[15] Consider only in extreme cases in which preferential treatment is evident; blinding is generally not expected in social interventions.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
Appraisal indicators Consider whether:						
1) *Outcome data are reasonably complete (80% or above).*[16]						
2) *If level of attrition (or other forms of missing/incomplete data) is more than 20%, are reasons for the missing data reported?*						
3) *If the level of attrition (or other forms of missing/incomplete data) is more than 20%, do the authors demonstrate similarity between remaining participants and those lost to attrition and are the proportion of participants with missing/incomplete data and reasons for missing/incomplete data similar across groups?*						
4) *If the level of attrition (or other forms of missing/incomplete data) is more than 20%, were appropriate statistical methods used to account for missing data? (e.g., sensitivity analysis)*[17]						
5) *If it is not possible to control for missing/incomplete data, are outcomes with missing/incomplete data excluded from analysis?*						
Low risk of bias	Moderate risk of bias	High risk of bias	Critical risk of bias	Worth continuing? Y/N		
VI. Bias in measurement of outcomes *Are measurements appropriate, e.g., clear origin or validity known?* **Appraisal indicators** Consider whether:						
1) *There was an adequate period for follow-up.*[18]						
2) *The outcome measure (e.g., employment status, income) was clearly defined and objective.*[19]						
3) *Outcomes were assessed using standardized instruments and indicators.*						
4) *Outcome measurements reflect what the experiment set out to measure.*						
5) *The methods of outcome assessment were comparable across groups.*						

[16] The assumption here that the level of attrition (or other forms of missing/incomplete data) is sufficiently low to not require adjustment.
[17] Select 'no' if the study addresses missing/incomplete data through simple estimates of missing data and observations.
[18] In many social science interventions, follow-up is not required to coincide with the start of the treatment; further, longer periods of follow-up are often required to measure changes.
[19] Subjective measures (e.g. those based on self-report) are likely to have lower reliability and validity than objective measures.

METHODOLOGICAL APPRAISAL CRITERIA				RESPONSE		
				Yes	No	Comment
6) *Were outcome assessors aware of the intervention received by study participants?*[20]						
Low risk of bias	Moderate risk of bias	High risk of bias	Critical risk of bias	Worth continuing? Y/N		
VII. Bias in selection of results reported *Are the reported outcomes consistent with the proposed outcomes at the protocol stage?* **Appraisal indicators** Consider whether:						
1) *It is unlikely that the reported effect estimate has been selected for publication due to it being a particularly notable finding among numerous exploratory analyses.*						
2) *It is unlikely that the reported effect estimate is prone to selective reporting from among multiple outcome measurements within the outcome domain.*						
3) *It is unlikely that the reported effect estimate is prone to selective reporting from among multiple analyses of the outcome measurements, including subgroup analysis.*						
4) *If subgroup/ancillary/adjusted analyses are presented, are these pre-specified or exploratory?*						
5) *The analysis includes an intention to treat analysis. (If so, was this appropriate and were appropriate methods used to account for missing data?)*[21]						
6) *Do the authors report on all variables they aimed to study (as specified in their protocol or study aims/research questions)?*						
Low risk of bias	Moderate risk of bias	High risk of bias	Critical risk of bias			

[20]Consider only in extreme cases in which preferential treatment is clearly evident; blinding is generally not expected in social interventions.
[21] Consider only in extreme cases in which preferential treatment is clearly evident; blinding is generally not expected in social interventions.

Appendix 4. APPROPRIATE FORMULA FOR EFFECT SIZE CALCULATIONS

Details of the appropriate formula for effect size calculations in reference to, and dependent on, the data provided in included studies are described below.

Studies reporting means (X) and pooled standard deviation for treatment (T) and control or comparison (C) at follow-up only:

$$d = \frac{x_{Tp+1} - x_{Cp+1}}{SD}$$

If the study does not report the pooled standard deviation, it is possible to calculate it using the following formula:

$$SD_{p+1} = \sqrt{\frac{(n_{Tp+1} - 1)SD_{Tp+1}^2 + (n_{Cp+1} - 1)SD_{Cp+1}^2}{n_{Tp+1} + n_{Cp+1} - 2}}$$

Where the intervention is expected to change the standard deviation of the outcome variable, we will use the standard deviation of the control group only.

For studies reporting means ($\underline{X}$) and standard deviations for treatment and control or comparison groups at baseline (p) and follow-up (p+1):

$$d = \frac{\Delta\underline{X}_{p+1} - \Delta\underline{X}_{p}}{SD_{p+1}}$$

For studies reporting mean differences ($\Delta\underline{X}$) between treatment and control and standard deviation (SD) at follow-up (p+1):

$$d = \frac{\Delta\underline{X}_{p+1}}{SD_{p+1}} = \frac{\underline{X}_{Tp+1} - \underline{X}_{Cp+1}}{SD_{p+1}}$$

For studies reporting mean differences between treatment and control, SE and sample size (n):

$$d = \frac{\Delta\underline{X}_{p+1}}{SE\sqrt{n}}$$

As primary studies have become increasingly complex, it has become commonplace for authors to extract partial effect sizes (e.g. a regression coefficient adjusted for covariates) in the context of meta-analysis. For studies reporting regression results, we will follow the approach suggested by Keef and Roberts (2004) using the regression coefficient and the pooled standard deviation of the outcome. Where the pooled standard deviation of the outcome is unavailable, we will utilize regression coefficients and SE or *t*-statistics to do the following, where sample size information is available in each group:

$$d = t\sqrt{\frac{1}{n_T} + \frac{1}{n_C}}$$

where *n* denotes the sample size of the treatment group and control.

We will use the following where only the total sample size information (*N*) is available, as suggested in Polanin and others (2016):

$$d = \frac{2t}{\sqrt{N}} \; Var_d = \frac{4}{N} + \frac{d^2}{4N}$$

We will calculate the *t*-statistic (*t*) by dividing the coefficient by the SE. If the authors only report confidence intervals and no SE, we will calculate the SE from the confidence intervals. If the study does not report the SE but reports *t*, we will extract and use this as reported by the authors. In cases in which significance levels are reported rather than *t* or SE (b), then *t* will be imputed as follows:

Prob > 0.1: $t = 0.5$

$0.1 \geq$ Prob > 0.05: $t = 1.8$

$0.05 \geq$ Prob > 0.01: $t = 2.4$

$0.01 \geq$ Prob: $t = 2.8$

Where outcomes are reported in proportions of individuals, we will calculate the Cox-transformed log odds ratio effect size (Sánchez-Meca and others, 2003):

$$d = \frac{ln\ (OR)}{1.65}$$

where OR is the odds ratio calculated from the two-by-two frequency table.

Where outcomes are reported based on proportions of events or days, we will use the standardized proportion difference effect size:

$$d = \frac{p_T\ -\ p_C}{SD(p)}$$

where p_t is the proportion in the treatment group and p_c is the proportion in the comparison group, and the denominator is given by:

$$SD(p) = \sqrt{p\ (1-p)}$$

where p is the weighted average of p_c and p_t:

$$p = \frac{n_T\ p_T\ +\ n_C\ p_C}{n_T\ +\ n_C}$$

An independent reviewer will evaluate a random selection of 10 per cent of effect sizes to ensure that the correct formulae will be employed in effect size calculations. In all cases after synthesis, we will convert the pooled effect sizes to commonly used metrics such as percentage changes and mean differences in outcome metrics typically used (e.g. weight in kg) whenever feasible.

Appendix 5. GRADING OF RECOMMENDATIONS ASSESSMENT, DEVELOPMENT AND EVALUATION (GRADE) TOOL (EXAMPLE)

CERTAINTY ASSESSMENT						SAMPLE SIZE	EFFECT	CERTAINTY	IMPORTANCE
Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Absolute (95% CI)		
Outcome 1									
RCT	Serious	Serious	Not serious	Serious	None	737	SMD **0.02 SD higher** (0.12 lower to 0.16 higher)	⊕◯◯◯ VERY LOW	Limited importance
Outcome 2									
RCT – 3	Serious	Serious	Not serious	Not serious	None	4,991	SMD **0.14 SD higher** (0.01 higher to 0.28 higher)	⊕⊕◯◯ LOW	Important, but not critical
Outcome 3									
RCT – 6 QED – 2	Very serious	Not serious	Not serious	Not serious	None	9,970	SMD **0.09 SD higher** (0.02 higher to 0.16 higher)	⊕⊕◯◯ LOW	Important, but not critical
Outcome 4									
RCT	Very serious	Serious	Not serious	Not serious	None	3,219	Two negative and three positive effect estimates with a 95% CI range of -0.08 to 0.16	⊕◯◯◯ VERY LOW	Important, but not critical
Outcome 5									
RCT	Very serious	Serious	Not serious	Not serious	None	3,219	SMD **0.02 SD higher** (0.09 lower to 0.05 higher)	⊕◯◯◯ VERY LOW	Important, but not critical
Outcome 6									
RCT	Not serious	Serious	Not serious	Serious	None	3,543	Five positive effect estimates with a 95% CI range of -0.00 to 0.41	⊕⊕◯◯ LOW	Important, but not critical

Certainty assessment						Sample size	Effect	Certainty	Importance
Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Absolute (95% CI)		
Outcome 7									
RCT - 7	Serious	Serious	Not serious	Not serious	None	8,359	SMD **0.06 SD higher** (0.02 lower to 0.14 higher)	⊕⊕◯◯ LOW	Critical
Outcome 8									
RCT – 2 QED – 1	Very serious	Serious	Not serious	Not serious	None	5,233	SMD **0.14 SD higher** (0.02 higher to 0.26 higher)	⊕◯◯◯ VERY LOW	Limited importance

Independent Evaluation Unit
Green Climate Fund
175, Art center-daero, Yeonsu-gu
Incheon 22004, Republic of Korea
Tel. (+82) 032-458-6450
ieu@gcfund.org
https://ieu.greenclimate.fund

International Fund for Agricultural Development
Strategy and Knowledge Department
Via Paolo di Dono, 44
00142 Roma, Italy
Tel.: (+39) 0654591
ifadknowledge@ifad.org
https://www.ifad.org