IEU LEARNING PAPER

07/2022

BEHAVIOURAL SCIENCE INTERVENTIONS WITHIN THE DEVELOPMENT AND ENVIRONMENTAL FIELDS IN DEVELOPING COUNTRIES: A SYSTEMATIC REVIEW

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

GREEN CLIMATE FUND | Independent Evaluation Unit

IFAD

GREEN CLIMATE FUND
INDEPENDENT EVALUATION UNIT

INTERNATIONAL FUND FOR
AGRICULTURAL DEVELOPMENT

Behavioural science interventions within the development and environmental fields in developing countries: A systematic review

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

07/2022

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

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

First Edition

This paper is a product of the Independent Evaluation Unit (IEU) at the Green Climate Fund (GCF). It is part of a larger effort to provide open access to the IEU's research and work and to make a contribution to climate change discussions around the world.

While the IEU has undertaken every effort to ensure the data in this report is accurate, it is the reader's responsibility to determine if any and all information provided by the IEU is correct and verified. Neither the author(s) of this document nor anyone connected with the IEU or the GCF can be held responsible for how the information herein is used.

Citation
The detailed citation for this paper is:
Ategeka, John, and others (2022). Behavioural science interventions within the development and environmental fields in developing countries: A systematic review. Learning paper (July). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund.

Credits
Head of the GCF Independent Evaluation Unit a.i.: Mr. Andreas Reumann
Task manager: Dr. Martin Prowse, Evaluation Specialist, Independent Evaluation Unit
Editing: Toby Pearce
Layout and design: Giang Pham
Cover photo: A laboratory activity to extract argan oil for commercial use in a project to reserve argan biosphere in Morocco funded by the GCF, ©Little Big Films /Green Climate Fund

A FREE PUBLICATION

Printed on eco-friendly paper

About the IEU

The IEU was established by the GCF Board as an independent unit, to provide objective assessments of the results of the Fund, including its funded activities, its effectiveness and its efficiency. The IEU fulfils this mandate through four main activities:

Evaluation: Undertakes independent evaluations at different levels to inform the GCF's strategic result areas and ensure its accountability.

Learning and communication: Ensures high-quality evidence and recommendations from independent evaluations are synthesized and incorporated into the GCF's functioning and processes.

Advisory and capacity support: Advises the GCF Board and its stakeholders of lessons learned from evaluations and high-quality evaluative evidence, and provides guidance and capacity support to implementing entities of the GCF and their evaluation offices.

Engagement: Engages with independent evaluation offices of accredited entities and other GCF stakeholders.

About the International Fund for Agricultural Development (IFAD)

IFAD is an international financial institution and specialized United Nations agency dedicated exclusively to transforming agriculture, rural economies, and food systems. It invests in poor rural people, empowering them to increase their food security, improve the nutrition of their families and increase their incomes. As a catalyst for increasing public and private investments in agriculture and the development of rural enterprises, IFAD helps small-scale producers build resilience, expand their businesses and take charge of their own development.

About the IEU's Learning Paper series

The IEU's Learning Paper series is part of a larger effort to provide open access to the IEU's work and to contribute to global discussion on climate change. The overall aim of the series is to contribute to learning and to add to global knowledge on what works, for whom, why, how much and under what circumstances, in climate change action. The findings, interpretations and conclusions are entirely those of the authors. They do not necessarily reflect the views of the IEU, the GCF or its affiliated organizations or of the governments associated with it. Comments are welcome and should be sent to ieu@gcfund.org.

About this IEU Learning Paper

This paper presents a systematic review on behavioural science interventions in the human development and environmental/climate fields in developing countries. The review synthesizes evidence on five specific behavioural interventions – feedback, reminders, salience (communication), salience (experience design), and goal setting – and how these influence development and environmental/climate outcomes. It finds that feedback and reminders show positive effects across contexts.

ACKNOWLEDGEMENTS

The authors are grateful for the comments received from the engagement committee for this evidence review, which has informed the choice of cells for the systematic review. The engagement committee for this evidence review includes Chaning Jang (Busara), Nathaniel Peterson (Busara), Erik Thulin (Rare) and Katie Williamson (Rare), ordered alphabetically by last name. The authors would also like to thank Junior Abdul-Wahab (GCF/IEU) for his comments on this systematic review. Any errors or inconsistencies are entirely the responsibility of the authors.

List of authors

Full name	Affiliation
John Ategeka	Africa Centre for Evidence, University of Johannesburg
Samantha Booth	Human Friendly Solutions
Romina Cavatassi	International Fund for Agricultural Development
Benjamin Curtis	Human Friendly Solutions
Viviane Filippi	International Fund for Agricultural Development
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 (led the overall evidence review project including the design and implementation of all research activities)
Martin Prowse	Independent Evaluation Unit, Green Climate Fund (led and coordinated the evidence review inputs from the GCF)
Jyotsna Puri	International Fund for Agricultural Development (led and coordinated the evidence review inputs from IFAD)
Jamie Robertsen	Independent Monitoring, Evaluation and Learning Consultant

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
CFL	Compact fluorescent lamp light
CI	Confidence interval
EGM	Evidence gap map
GCF	Green Climate Fund
GRADE	Grading of recommendations, assessment, development and evaluations
IEU	Independent Evaluation Unit
IFAD	International Fund for Agricultural Development
PICOS	Population, intervention, comparison, outcome, study design
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
RCT	Randomized controlled trials
REML	Random effects restricted maximum likelihood
SD	Standard deviation
SE	Standard error
SMD	Standardized mean difference
SR	Systematic review
UNEP	United Nations Environment Programme
WASH	Water, sanitation and hygiene

EXECUTIVE SUMMARY

Climate change is projected to intensify over the coming decades, resulting in dramatic impacts on natural and human systems. This is a pressing problem given that the impacts of climate change will not be uniformly distributed across the globe. Developing countries are likely to be disproportionately affected due not only to their exposure to shocks and stresses but also their limited capacity to withstand and respond to damaging fluctuations. Human behaviour is a key driver of climate change, and behavioural science interventions offer promising opportunities to promote positive environmental/climate and development outcomes. Yet rigorous empirical guidance is lacking in terms of how to change behaviour most effectively to support adaptation and emissions reductions in developing countries.

Objectives

The primary objective of this systematic review (SR) is to identify, assess and synthesize evidence on the effectiveness of feedback, reminders, salience (communication), salience (experience design), and goal-setting interventions conducted in developing countries on environmental/climate and development outcomes. It aims to facilitate the use of evidence in informing policy and practice decisions within the environmental/climate and development fields, particularly in the GCF and IFAD.

Methods

We first conducted an interactive evidence gap map (EGM) using a broader research scope to provide an overview of the evidence base on the effectiveness of behavioural science interventions, covering 22 interventions. From a landscape of 34,340 papers, this map included a total of 84 studies.

In consultation with the advisory group, the SR focused on only five behavioural science interventions: feedback, reminders, salience (communication), salience (experience design), and goal-setting interventions. The full SR followed an aggregative review approach, including only the rigorous quantitative impact evaluations of these five interventions. To identify relevant research studies, we conducted an exhaustive search of 68 academic and grey literature sources. Screening led to the inclusion of 60 impact evaluations of interventions across the five intervention types. The 60 included studies were subject to a detailed data extraction and critical appraisal process. We used statistical meta-analysis to establish the overall effects of different interventions, and narrative synthesis where the identified evidence base did not allow us to statistically pool studies' results.

Selection criteria

We included randomized controlled trials (RCTs) as well as quasi-experimental studies that used quantitative methods. Studies could be either published, peer-reviewed articles or grey literature such as working papers, reports and dissertations. Studies must have been published on or after 2000, employed a behavioural science intervention and include at least one of our outcomes of interest: knowledge, uptake and use, behavioural outcomes, development results and impact. Studies had to be conducted in developing countries to be included.

Results

We conducted 12 meta-analyses in our SR to examine the overall effects of feedback, reminders and goal-setting interventions aiming to improving knowledge, uptake and use, behavioural outcomes and development results. The most effective intervention type according to this analysis is feedback, particularly in relation to behavioural outcomes of electricity and water consumption. We identified

an overall pooled effect estimate of 0.26 (CI: 0.13 to 0.39), which is assessed as being moderate quality of evidence using the GRADE framework. A similar pattern emerges for reminders, specifically on acquisition of knowledge, where we identify an overall pooled effect estimate of 0.87 (CI: 0.34 to 1.41), which we also find to be comparatively effective. However, this finding is based on a lower quality of evidence than for the feedback intervention according to the GRADE framework. We find no significant effects on goal-setting interventions based on a limited number of meta-analyses. For salience (experience design) and salience (communication) interventions identified in this SR, the heterogeneity in interventions and outcomes does not allow for a meta-analysis and rigorous synthesis of effects.

Authors' conclusions

We identified an evidence base that is heterogeneous in terms of the applied interventions and overall of low quality in terms of the methodological trustworthiness of studies and consistency of effects. Given the limitations of this evidence base, our review conducted only limited meta-analyses in three of the intervention types, namely feedback, reminders and goal-setting interventions. Our meta-analysis provided cautious evidence that interventions that deploy feedback and reminders are an effective behavioural approach to improving human and development outcomes in developing countries. Feedback and reminders as tools for influencing behaviour should receive particular consideration by the GCF and IFAD. For the remaining three behavioural science interventions in the review – goal setting, salience of communication and salience of experience design – more research is required to inform decision-making. The very low number of studies and evidence available in the field and its scattered nature in terms of types of interventions, strengthens the importance of this work regarding the need to generate and use more of this type of evidence.

I. INTRODUCTION

A. THE PROBLEM, CONDITION OR ISSUE

The already on-going process of climatic change is projected to intensify over the next several decades, resulting in dramatic impacts on natural, human and productive systems. The need for both adaptation and mitigation is clear. Equally clear is that human behaviour is a key driver of climate change and that many adaptation and mitigation strategies require behaviour changes. However, rigorous empirical guidance is lacking in terms of how to change behaviour most effectively to support adaptation and emissions reductions. In particular, research evidence relating to developing countries is somewhat thin and scattered. This is a pressing problem given that the negative impacts of climate change will not be uniformly distributed across the globe; developing countries are likely to be disproportionately affected due not only to their exposure to shocks and stresses but limited capacity to withstand and respond to damaging fluctuations (see Global Commission on Adaptation, 2019; Solomon and others, 2007; Edenhofer and others, 2014; UNEP, 2017; Wade and Jennings, 2015; Binet and others, 2021). Therefore, understanding what is effective in changing behaviour in these countries is important. Greenhouse gas emissions from human behaviours, such as the burning of fossil fuels, 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 behaviour to reduce carbon emissions (Williamson and others, 2018). Hence, climate change may be mitigated by altering behaviours related to these sectors.

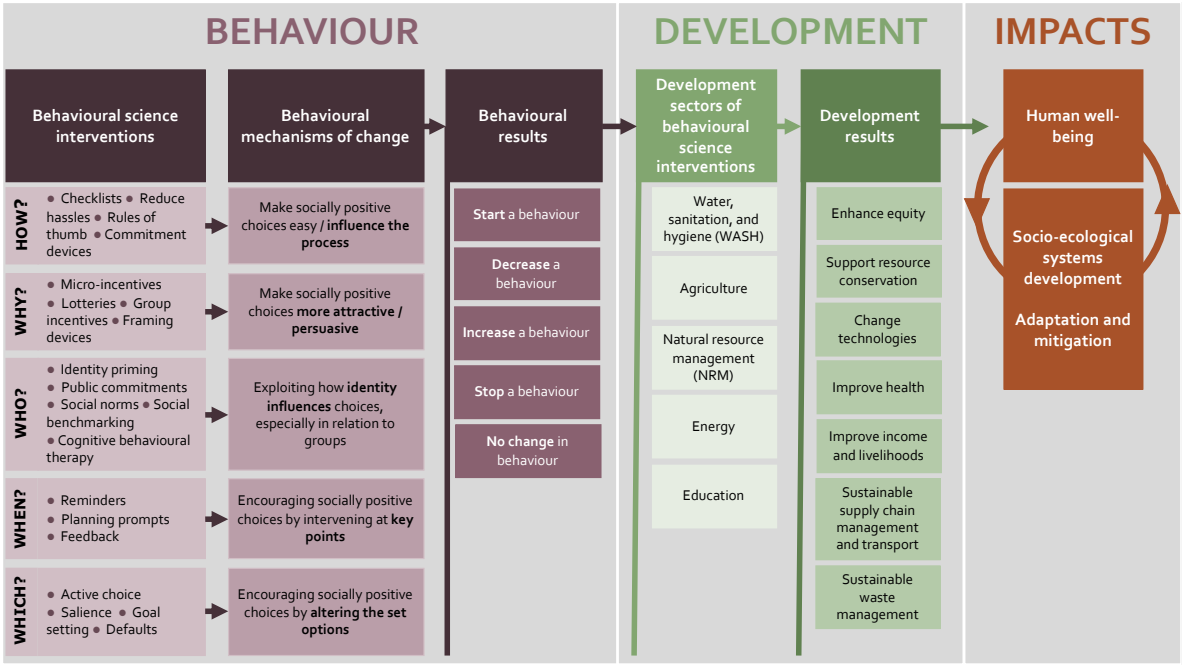
In recent decades, theories and evidence from behavioural science – defined by Balmford and others (2021) as the scientific study of behaviour informed by an array of disciplines including sociology, psychology, economics, anthropology, and political science – have provided insights into the social, motivational, cognitive, cultural, and contextual factors underlying human behaviour. Stern (2020) describes behavioural interventions as involving neither command and control regulations nor financial incentives and as methods to change behaviour, without changing pay-offs from action. Examples include information provisions, appeals to values and norms, or engagement and restructuring choice options (commonly referred to as nudges). These insights have informed interventions that have helped to encourage societally-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 a variety of 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). This review presents the results of a systematic review (SR) of evidence on the effectiveness of behavioural science interventions, namely feedback, reminders, salience (communication), salience (experience design), and goal setting on environmental/climate and development outcomes in developing countries, with particular focus on data collection and analysis.

B. THE INTERVENTION AND HOW THE INTERVENTION MIGHT WORK

The intervention logic and how the behavioural science intervention might lead to desired effects is outlined in a theory of change. A theory of change is essentially “a set of statements that describe

the process and the mechanisms (i.e. the how and why)" through which an intervention is thought to work and the results it aims to affect (Jones and Rosenberg, 2018). In the context of the evidence review on behavioural change, the purpose of the theory of change is to inform the types of interventions included in the EGM. The theory of change directly informed the Population, Intervention, Comparison, Outcome and Study (PICOS) design framework that was used to develop inclusion and exclusion criteria. It also illustrates the role that behavioural science interventions play in human and environment and development outcomes. The theory of change is shown in Figure 1.

Figure 1. Theory of change for this evidence review



Source: Authors

The theory of change is divided into three distinct parts: behaviour, development and impact. Behaviour outlines three levels, while development provides two levels, with impact as the culmination of both behaviour and development interventions. The theory of change conceptualizes three levels showing how behavioural interventions lead to behavioural results, through five behavioural mechanisms of change. Definitions of the behavioural interventions are provided in Table 1 below. The first level is a categorization of different behavioural interventions (i.e. checklists, social norms or defaults). These interventions are most commonly applied in the field and are drawn from the list compiled by the Behavioural Evidence Hub, a leading knowledge clearinghouse for policy-relevant behavioural science. The second level specifies the mechanisms of change – for example, how these interventions actually influence behaviour such as through changing sets of options or "nudging" at key decision points. These mechanisms are informed by two prominent conceptualizations of behaviour change: the EAST framework produced by the Behavioural Insights Team (Service and others, 2014) and the 4Ps Framework for Behaviour Change from Yale University (Chance and others, 2014). The third level outlines concrete behavioural results (e.g. starting a behaviour, stopping a behaviour).

This theory of change is unique in that the outcome of the behavioural intervention leads to activities that provide inputs for the development component of the theory of change. There are

therefore two intervention levels before results are attained in human and environmental development. These development results, which are grouped by indicative sectors, then have their own intermediate and wider outcomes. The development results are purposefully categorized more broadly than the behavioural change interventions and results. This is to ensure that (1) the theory of change is not so complex as to lose utility and (2) the theory of change does not limit the development results in the evidence gap mapping process. The transition from narrowly defined interventions to broad development results also means that the causal pathways are less well-articulated. Two examples that outline potential causal pathways are provided below:

- If the development result was to adopt new farming practices (to improve income and livelihoods) through an agriculture intervention, the potential causal pathway would be using planning prompts (behavioural interventions) to encourage socially positive choices by intervening at key decision points (behavioural mechanism of change). This would then result in starting a behaviour (behavioural result), which in this example would be adopting new farming practices.
- If the development result was to use more energy-efficient lighting (to change technologies), through an energy-related intervention, the potential causal pathway would be using micro-incentives (behavioural interventions) to make positive choices more attractive/persuasive (behavioural mechanism of change). This would then result in starting a behaviour (behavioural result), which in this example would be using energy-efficient lighting.

The development sectors were selected based on their potential for behavioural interventions with outcomes that have results impacting socioecological systems. Since the theory of change is a living document, it can reflect the ongoing findings of the review. Similarly, the development results were selected as part of an iterative process of refinement. They are not necessarily mutually exclusive, and one intervention could target more than one result area.

The impact level denotes the desired state of stable socioecological systems through human well-being and climate change adaptation and mitigation. These two impacts are intrinsically linked. For the purposes of this study, we considered the intention of the research when determining contribution towards impact. Using the causal pathway examples provided above, examples of the impact level could include the following:

- Adjusting farming practices to new climate conditions contributes to climate change adaptation and improves human well-being through sustaining or improving incomes and livelihoods. This in turn contributes to developing and sustaining more stable socioecological systems.
- Changing technologies by using energy-efficient lighting contributes directly to mitigating the effects of climate change by reducing energy consumption. This in turn contributes to developing and sustaining more stable socioecological systems.

C. WHY IT IS IMPORTANT TO DO THIS SR?

This review examines the effectiveness of behavioural science interventions, namely feedback, reminders, salience and goal setting in promoting environmental and development outcomes by individuals, households, communities, and companies in developing countries. The focus of the review is due to the growing hopes for behavioural interventions (Schot and others, 2016; Stern and others, 2016) such as consumption feedback, social comparison messages, or tailored environmental appeals as potentially cost-effective strategies compared to traditional market tools and regulation.

To the best of our knowledge, there is no SR evidence that carefully explores the effectiveness of these behavioural science interventions on environmental and development outcomes in developing

countries. While there is extensive evidence about both what is ineffective, and what works in promoting behaviour change broadly (Flanagan and Tanner, 2016), this evidence base has not been rigorously synthesized in the climate sector in developing countries specifically. This review reduces this gap within the literature to inform governments, donors, development practitioners and other decision-makers about the available evidence on a broad set of behavioural science interventions and their impacts on climate change mitigation and adaptation (including human development) across different sectors in developing country contexts. Through this review, we rigorously synthesize evidence to discern whether feedback, reminders, salience (communication), salience (experience design), and goal-setting interventions work to attain desired environmental and development outcomes in developing countries. These interventions are described in detail in Appendix 1.

D. OBJECTIVES

This SR aims to facilitate the use of evidence in informing policy and practice decisions within the environmental and development fields, particularly climate change 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 publication status, evaluation design and context?

To what extent do implementation features moderate the effectiveness of these behavioural science intervention programmes?

II. METHODS

We use a two-stage SR approach; the first stage consists of an EGM, which has been completed,[1] and the second stage is an SR and synthesis, in compliance with guidelines for the production of EGMs and SRs outlined by the Campbell Collaboration.[2] We adopt a transparent and adaptable research process that fully integrates the conduct of the EGM and the SR. Suggestions based on previous synthesis project experiences in the environmental sector focused on the payment for ecosystem services (Snilstveit and others, 2019) and the gender sector (Langer and others, 2018), indicate that the successful integration of an evidence map and subsequent full SR 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 of rigour 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 backend as well as the frontend of the evidence mapping tool applied to integrate the knowledge management aspect of the evidence review, with the visualization requirements of the EGM.

A. EVIDENCE GAP MAP

The EGM had a broader scope in terms of inclusion of evidence than the full SR, but both are focused on the nature of the evidence that exists regarding the effectiveness of behavioural science interventions on environmental/climate or development outcomes in developing countries. The EGM mapped evidence from impact evaluations and SRs across 22 behavioural science interventions. Its main objective was to indicate the overall nature and size of the available evidence base, identify areas for synthesis and substantiate evidence gaps for future commissioning. The SR will focus on five of the 22 selected interventions, namely feedback, reminders, salience (communication), salience (experience design), and goal setting.

The EGM is a product in its own right that supports stakeholder engagement with the evidence base and also supports decision-making about the most effective synthesis approach and scope. The final map presented to GCF and IFAD included a total of 84 studies (82 impact evaluations and two SRs). The EGM was applied instrumentally to guide discussions about which areas of the evidence base to use for synthesis, as well as what the most effective method to synthesize the evidence to answer the review question would be. Following two meetings with the advisory group to jointly identify the relevant areas of evidence for the SR, the synthesis gaps and stakeholders' interest steered the focus of the review to the five intervention types.

[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. The approach paper and technical report are available at https://ieu.greenclimate.fund/evidence-review/behavioural-science.
[2] Systematic reviews are available at https://onlinelibrary.wiley.com/pb-assets/Campbell%20Policies%20and%20Guidelines%20Dec2020-1608292090217.pdf. Evidence and gap maps are available at https://onlinelibrary.wiley.com/doi/10.1002/cl2.1125.

B. SYSTEMATIC REVIEW AND SYNTHESIS

A total of 68 unique studies were identified from the EGM that focus on these five intervention types. An effectiveness review was conducted to answer the review question about the extent to which these interventions have been effective at achieving behavioural change in desired environmental and development outcome areas in developing countries. Therefore, the SR only includes primary studies that measure the effects of interventions and have designs that can reliably attribute observed effects to these applied interventions. Individual effects are synthesized into overall estimates of treatment effects using statistical meta-analysis.

a. Intervention-outcome framework for the EGM

The EGM intervention-outcome framework was used to structure and visualize the evidence-base, and its design is directly influenced by the theory of change. Comprehensive details of the intervention-outcome framework are provided in the approach paper (Booth and others, 2022a) and the EGM report (Booth and others, 2022b).

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

To systematically synthesize literature on the effectiveness of the five intervention types, an underlying focus on environment and human development outcomes guides the scope of the review. Formally, we adopt the PICOS framework to develop our inclusion criteria. Full details of the inclusion criteria for the SR are highlighted in the approach paper (Booth and others, 2022a) and the SR protocol (Booth and others, 2022c). The inclusion criteria define the precise characteristics of the studies that are included in the SR. All evidence not meeting these criteria was excluded from this review.

c. Searching for evidence

A comprehensive search strategy was developed to search for qualifying research literature studies to identify all available evidence relevant to the review question, and will be included in the SR. The approach paper and EGM report outlines the search strategy, including sources (databases and repositories), backward and forward searches, search terms, combination of search terms, and results from the searching and screening process.

III. DATA COLLECTION AND ANALYSIS

A. SELECTION OF STUDIES

Once we obtained the search results, they were imported into the SR software EPPI-Reviewer 4.[3] This platform is used to manage references, identify and remove duplicate studies, and screen records for inclusion using the procedures outlined below. This review management software (EPPI-Reviewer 4) was used to manage the entire review process. Search results from organizational websites and the citation searches were captured in MS Word, and only studies deemed to be relevant for the EGM were transferred to EPPI-Reviewer 4. Studies that were not already on EPPI-Reviewer were captured manually on the software. Before proceeding with screening, all duplicate titles were excluded from the review using the duplicate control function on EPPI-Reviewer 4.

At the title and abstract screening level, we conducted a manual double-screening exercise to assess the eligibility of studies using the inclusion criteria highlighted above, and decisions made about each citation were recorded on the same platform. To ensure quality and consistency in the screening process, 5,000 studies were double screened at the title and abstract level. Two reviewers screened this common sample of 15 per cent of all study abstracts. During the training, the results given by the researchers were compared, and any discrepancies in coding decisions were discussed as needed, including clarification of the inclusion criteria. The individual screening was only permissible once a similarity index of the screening exercise reached 90 per cent.

We conducted full-text screening of each study that met all title and abstract screening inclusion criteria. Two reviewers independently examined the full text of each study in detail against the protocol and independently decided whether to include or exclude the study. Any disagreements between reviewers were reconciled through the supervision of a senior review team member. The output of this stage is a set of studies deemed suitable for inclusion in the EGM.

The EGM had a broader scope in terms of inclusion of evidence than this SR, but both are focused on the nature of the evidence that exists regarding the effectiveness of behavioural science interventions on climate, environmental and development outcomes in developing countries. The EGM mapped evidence from impact evaluations and SRs across 22 behavioural science interventions. The final map presented to GCF and IFAD included a total of 84 studies (82 impact evaluations and two SRs). The EGM was applied instrumentally to guide discussions about which areas of the evidence base to use for synthesis, as well as what the most effective method would be to synthesize the evidence to answer the review question. Following two meetings with the advisory group to jointly identify the relevant areas of evidence for the SR, the synthesis gaps and stakeholders' interest steered the focus of the review to feedback, reminders, salience (communication), salience (experience design), and goal setting interventions, constituting the focus of this review.

B. DATA EXTRACTION AND MANAGEMENT

We used a predefined data extraction tool to extract data systematically and transparently from the included primary studies. The coding tool highlighted in Appendix 2 was translated into Excel to

[3] EPPI-Reviewer 4 is software for all types of literature review, including SRs, meta-analyses, "narrative" reviews and meta-ethnographies. More information is available at https://eppi.ioe.ac.uk/CMS/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4&.

extract information that is required for the SR and synthesis. The data was entered directly into Excel, with full-text reports 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.[4]
- Methodological information, analysis method, and type of comparison.
- Quantitative data for outcome measures, including outcome descriptive information, sample size in each intervention group, outcomes means and standard deviations, and test statistics (e.g. t-test, F-test, p-values, 95 per cent confidence intervals [CI]).
- Information on intervention design, including how the intervention incorporates participation, participant adherence, contextual factors, and programme mechanisms including implementation fidelity.

C. Assessment of risk of bias in included studies

We applied a critical appraisal tool to assess the trustworthiness of the impact evaluations included in the SR. Trustworthiness refers to whether the confidence that the findings reported in the included impact evaluations were rigorous and credible and are likely to reflect the results of the evaluated interventions rather than reflecting the influence of the applied study design and research conduct. To assess the risk of bias of the primary studies, we adapted the Cochrane risk of bias tool for randomized and non-randomized studies (Sterne and others, 2016), shown in Appendix 3. We have previously used and adopted this risk of bias tool in international development reviews (Stewart and others, 2015; Rebelo Da Silva and others, 2017). Sterne and others (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 domain of bias received 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 were included in the review but excluded from the synthesis. The critical appraisal tool used to assess studies for the SR is presented in Appendix 3, and it was piloted using a similar approach to that used for piloting the data extraction tool. Two reviewers independently assessed each study, then came together to compare their decisions. Where those reviewers were in disagreement about the risk of bias rating for a particular study, a third reviewer was consulted.

D. Measures of treatment effect

1. 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 with dependent effect sizes might occur. For example, there could be several publications that stem from one study, or several studies based on the same data set. Some studies might have multiple treatment arms that are all 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. All such cases yield a set of statistically dependent effect size estimates (Borenstein and others, 2009).

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

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

We extracted effects reported across different outcomes or subgroups within a study, and 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 was extracted. Where studies reported effects from multiple model specifications, we adopted the author's preferred model specification. If this was not stated or was unclear, the specification with the most controls was used. Where studies reported multiple outcomes or evidence according to sub-groups of participants, we recorded and reported data on relevant sub-groups separately. Further information on criteria for determining independent effect sizes is presented below.

We dealt 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 used effect sizes from the most recent publication. For studies with outcome measures at different time points, we followed De La Rue and others (2014) and synthesise 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 existed within these periods, we adopted the most recent measure. When studies included multiple outcome measures to assess related outcome constructs, we followed Macdonald and others (2012) and selected the outcome that appears to reflect the construct of interest most accurately without reference to the results.

If studies included multiple treatment arms with only one control group and the treatments represent separate treatment constructs, we calculated the effect size for treatment A versus control and treatment B versus control, and included them in separate meta-analyses according to the treatment construct. If treatments A and B represented variations of the same treatment construct, we calculated the weighted mean and standard deviation for treatment A and B before calculating the effect size for the merged group versus control group, following the procedures outlined in Borenstein and others (2009, chapter 25). Where studies report data from two or more independent subgroups, we calculated the weighted mean and standard deviation for treatment A and B before calculating the effect size for the merged group versus control group, following the procedures outlined in Borenstein and others (2009, chapter 23). Where different studies report on the same programme but use different samples (e.g. from different regions or separately for men and women) we included both estimates, treating them as independent samples, provided that effect sizes are measured relative to separate control or comparison groups.

2. Effect size calculations

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 per cent CIs) was extracted using Excel. Effect size data was stored and any necessary cleaning was conducted in Excel. Following the screening and descriptive data extraction process of ensuring consistency in coding quality, two reviewers piloted the effect size data extraction tool, working independently on a random sample (10 per cent) of included studies to test the tool across a range of the included impact evaluation designs and methods. We aimed 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, the remaining studies were coded by individual reviewers and all data extracted was checked by a third reviewer.

An effect size expresses the magnitude (or strength) and direction of the relationship of interest (Valentine and others, 2015; Borenstein and others, 2009). We extracted 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 calculated the standardized mean difference (SMDs), or Cohen's *d*, its variance, and standard error using formulae provided in Borenstein and others (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):

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

We choose an appropriate formula for effect size calculations in reference to, and dependent on, the data provided in included studies. For example, for studies reporting means (X) and pooled standard deviation (SD) 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 SD, it is possible to calculate it using the following formula:[5]

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

Where the intervention is expected to change the SD of the outcome variable, we used the SD of the control group only.

For studies reporting means ($\underline{X}$) and standard deviations (SD) 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, standard error (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 followed the approach suggested by Keef and Roberts (2004) using the regression coefficient and the pooled SD of the outcome. Where the pooled SD of the outcome is unavailable, we utilized regression coefficients and standard errors 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.

[5] $n_{Tp+1} + n_{Cp+1}$ are sample sizes for treatment and control at follow-up.

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 calculated the *t*-statistic (*t*) by dividing the coefficient by the SE. If the authors only report CIs and no SE, we calculated the SE from the CIs. If the study does not report the SE but reports *t*, we extracted and used this as reported by the authors. In cases in which significance levels are reported rather than *t* or SE (b), then *t* we imputed as follows:

Prob > 0.1: $t = 0.5$

0.1 ≥ Prob > 0.05: $t = 1.8$

0.05 ≥ Prob > 0.01: $t = 2.4$

0.01 ≥ Prob: $t = 2.8$

Where outcomes are reported in proportions of individuals, we calculated 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 used 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 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 evaluated a random selection of 10 per cent of effect sizes to ensure that the correct formulae was employed in effect size calculations.

E. UNIT OF ANALYSIS ISSUES

Unit of analysis errors can arise when the unit of allocation of a treatment is different to the unit of analysis of effect size estimate, and this is not accounted for in the analysis (e.g. by clustering standard errors at the level of allocation). We assessed studies for unit of analysis errors (Campbell Collaboration, 2019), and where they existed, we corrected for them by adjusting the standard errors according to the following formula (Higgins and others, 2020; Waddington and others, 2012; Hedges 2009):

$$SE(d)' = SE(d) * \sqrt{1 \; + (\mathrm{m} - 1)\mathrm{c}}$$

where *m* is the average number of observations per cluster and *c* is the intra-cluster correlation coefficient. Where included studies used robust Huber-White standard errors to correct for clustering, we calculated the SE of *d* by dividing *d* by the t-statistic on the coefficient of interest.

F. ASSESSMENT OF HETEROGENEITY

To visibly examine variability in the effect size estimates, we used forest plots to display the estimated effect sizes from each study along with their 95 per cent CIs. Subsequently, and acknowledging the limitations of quantification of heterogeneity and the different strengths of statistical approaches, the following tests for heterogeneity were conducted: calculation of the Q-statistic as a statistical test of heterogeneity (Hedges and Olkin, 1985); and calculation of the I^2 and Tau^2 ($\hat{\tau}^2$) 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). In addition, we explored heterogeneity using moderator analysis in meta-regression specifications where there were at least four studies in each category and significant heterogeneity. It has been suggested that a minimum of 10 studies are required for moderator analysis (Borenstein and others, 2009), but there are no hard and fast rules. However, we ensured that for categorical moderators, there were a minimum of two studies in each category per moderator variable for meaningful moderator analysis.

G. ASSESSMENT OF REPORTING BIASES

When the dissemination of research findings is influenced by the nature and direction of results, this leads to reporting biases. Positive and statistically significant results that show an intervention works are not only more likely to be published but more likely to be published rapidly, usually in English. Further, there is a high probability they will be published more than once, in high-impact journals and therefore will be more likely to be cited by others. In SRs, the contribution made to the totality of the evidence by studies with non-significant results is the same as that of studies with statistically significant results. Funnel plots are used to test for reporting biases, but the general rule is that they are utilized only when there are at least 10 studies included in the meta-analysis. This is because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry (Higgins and others, 2019). To reduce the possibility of publication bias, we searched for and included unpublished studies in this review. Due to the limited number of studies in each meta-analysis, tests for publication bias were not feasible.

H. DATA SYNTHESIS

From the included studies in the EGM covering the five selected interventions of interest to the SR, we conducted meta-analyses of studies that we assess to be sufficiently similar. Most studies feature multi-components (combination of two or more interventions) and multiple treatments groups. The minimum criteria were to only combine studies using meta-analysis when we identify two or more effect sizes using a similar outcome construct and where the comparison group stated is judged to be similar across the effects, which is similar to the approach taken by Wilson and others (2011). We combined studies in the same analysis when they evaluate the same intervention type and the same type of outcome measure. We conducted separate analyses for the major outcome categories for each intervention group on knowledge uptake and use outcomes, behavioural outcomes, and development results. Our approach was to select the lowest level of outcomes in the pathway of change (theory of change indicated above).

We used the metan package in Stata 16 software to conduct the meta-analyses. The amount of heterogeneity (i.e., $\hat{\tau}^2$), was estimated using the DerSimonian-Laird estimator (DerSimonian and Laird, 1986). In addition to the estimate of $\hat{\tau}^2$, the Q-test for heterogeneity (Cochran, 1954) and the

I^2 statistic (Higgins and Thompson, 2002) are reported. Where there were too few studies or included studies are considered too heterogeneous in terms of interventions or outcomes, we presented a discussion of individual effect sizes along the causal chain through a narrative synthesis.

I. SUBGROUP ANALYSIS AND INVESTIGATION OF HETEROGENEITY

Whenever feasible, we conducted moderator analyses to explain variations in effect sizes. Moderators are variables such as socio-economic 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 (Gough and others, 2017), we utilized 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 vs. private sector vs. government investments), publication type, publication date.
- Methodological characteristics: study design, risk of bias, evaluation period, length of intervention.
- Substantive characteristics: participant characteristics (gender, age, socio-economic status), context (geographical setting), intervention type, intervention length, intervention features, type of implementing agency.

J. SENSITIVITY ANALYSIS

To test the robustness of the results of the meta-analysis, a number of sensitivity analyses were conducted. Broadly, this involves 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 anticipated that the included studies would vary methodologically and therefore conducted sensitivity analyses to examine the influence of these variations on the summary measures, in order to offer possible explanations for the differences between studies when interpreting the results. 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 study design and methodological decisions of the review team.

K. STRENGTH OF THE EVIDENCE ASSESSMENT

The last research step in the SR was 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, and considered additional factors to assess the overall body of evidence and how much trust can be placed in recommendations based on it. Appendix 4 presents the GRADE results for each meta-analysis conducted.

IV. RESULTS

In this section, we report descriptive results for the review to provide an overview of the characteristics and distribution of the evidence base across all interventions covered in the review. We start by providing the results of the search and screening of the literature followed by a section providing a summary of the characteristics of included studies.

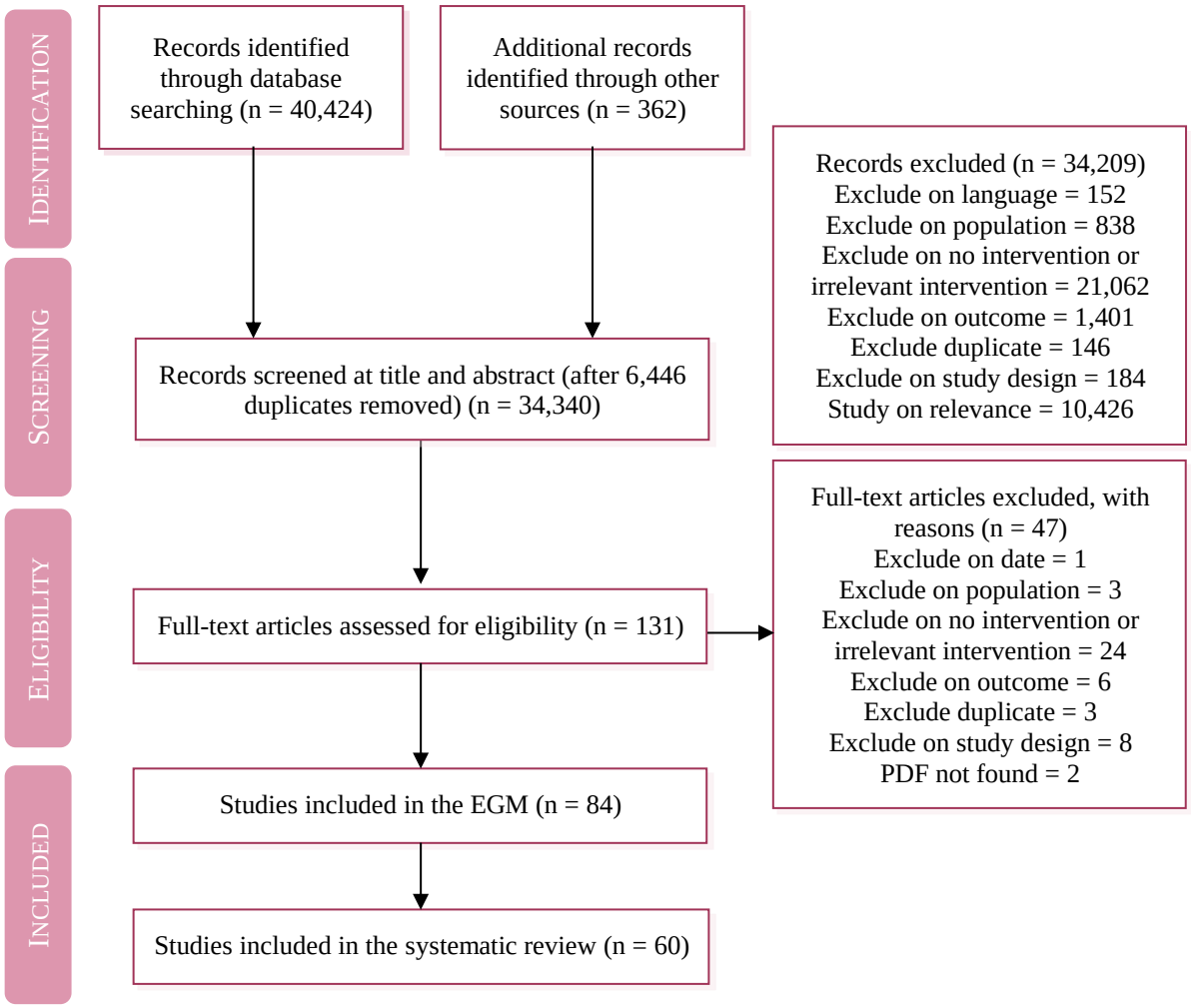
A. DESCRIPTION OF STUDIES: SEARCH RESULTS AND CHARACTERISTICS OF THE EVIDENCE BASE

1. RESULTS OF THE SEARCH

We conducted our search in January and February 2022. As the PRISMA diagram[6] (Moher and others, 2009) below shows, the search strategy returned 40,424 records (Figure 2). After removing duplicates, 34,340 records were left for screening at the title and abstract level. As noted above, to enable ex-post validation of screening consistency, approximately 15 per cent (5,000) of the 34,340 studies were retained for double screening at title and abstract. Around 90 per cent of these studies were screened consistently by all screeners. Consistency checks during closer reviews of the text and coding yielded similar consistency rates. Screening these records, we identified 131 studies to review at the full-text level. Of them, 24 were excluded because the interventions were not relevant to the scope of the EGM. A total of eight studies were excluded due to their study design; six were excluded due to irrelevant outcomes; three studies were excluded on population, and three studies were identified as duplicates. One study was excluded based on its publishing date and two full texts were not found. The final set of studies that made it into the evidence map comprised 84 studies, of which 82 were impact evaluations and two were SRs. Of these, 60 studies were included in the SR focusing on the interventions of interest noted above.

[6] PRISMA stands for preferred reporting items for SRs and meta-analyses. More information is available at http://prisma-statement.org/PRISMAStatement/PRISMAStatement.aspx.

Figure 2. PRISMA diagram



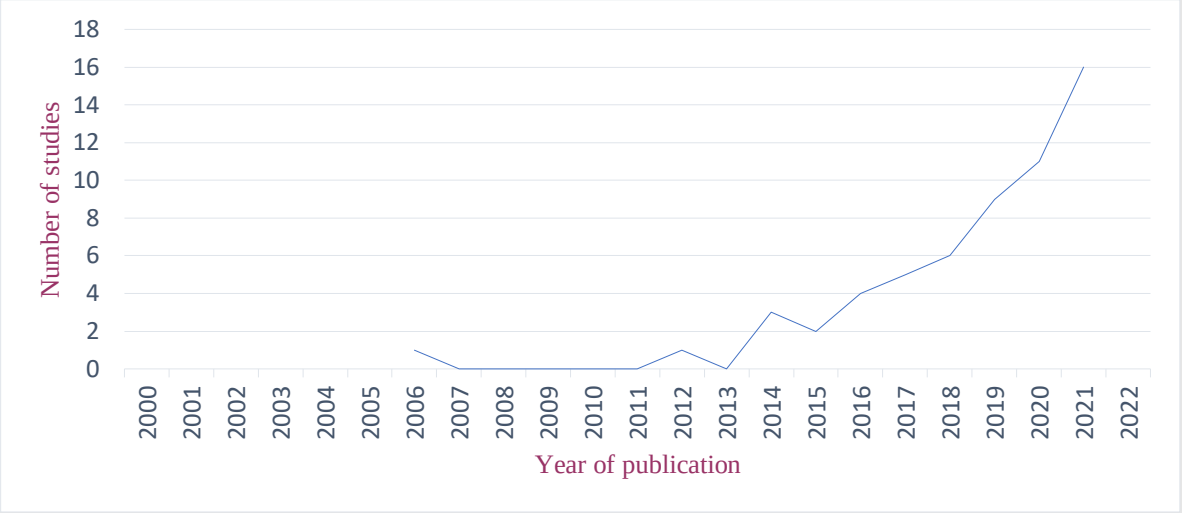
Source: Authors

2. Characteristics of included studies

a. Publication trend over time

Figure 3 reports the publication trend of the IEs included in the EGM over time. In line with this review's specific focus on interventions that commenced on or after 2000, the earliest included publication was Ashraf and others (2005) and the most recent was Klege and others (2022). The annual number of publications saw a steady increase over the years with 52 publications (87 per cent of the total) between 2016 and 2022. The year 2021 ranked the highest, with 16 publications.

Figure 3. Publication trend over time



Source: Authors

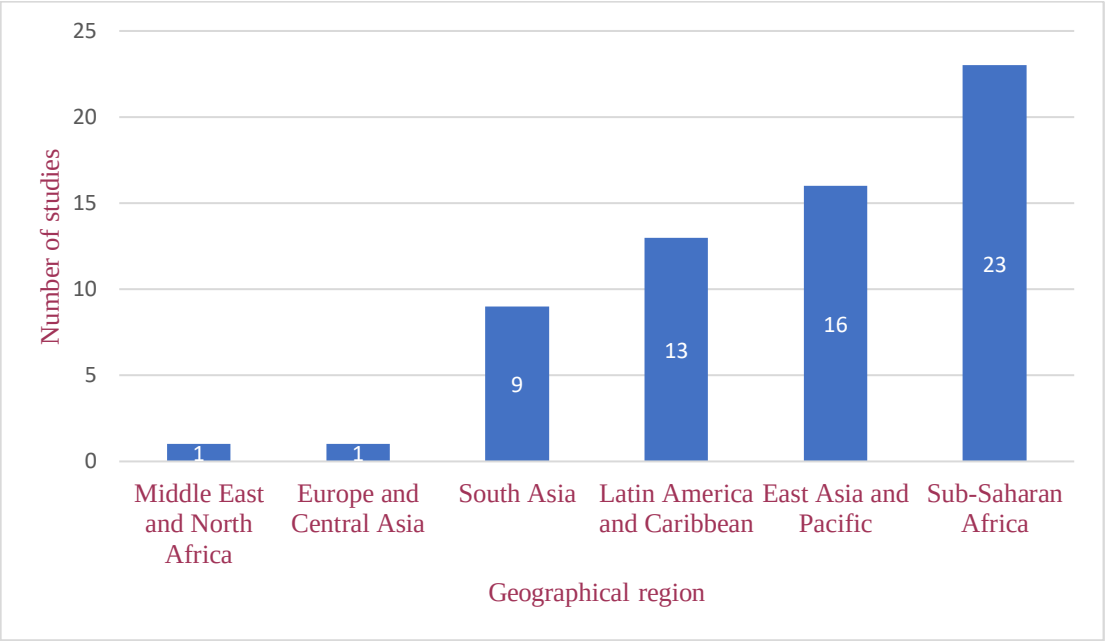
b. Geographic distribution

The included studies retained after the screening process were geographically diverse (see Figure 4 and Figure 5) and representative of developing contexts around the globe. The studies were conducted across 30 countries. Thirty-five per cent (23) of the studies were conducted in sub-Saharan Africa, 25 per cent (16) in East Asia and the Pacific, and 21 per cent (13) in Latin America and the Caribbean. Fifteen per cent were carried out in South Asia (9), and only one study was conducted in the Middle East and North Africa as well as one in Europe and Central Asia. As shown in Figure 6, most of these interventions were carried out in lower-middle-income countries (24; 40 per cent) and upper-middle-income countries (19; 32 per cent). The rest of the interventions were conducted in low-income countries (8; 14 per cent) and high-income countries (8; 14 per cent).

Figure 4. Geographic spread of included studies

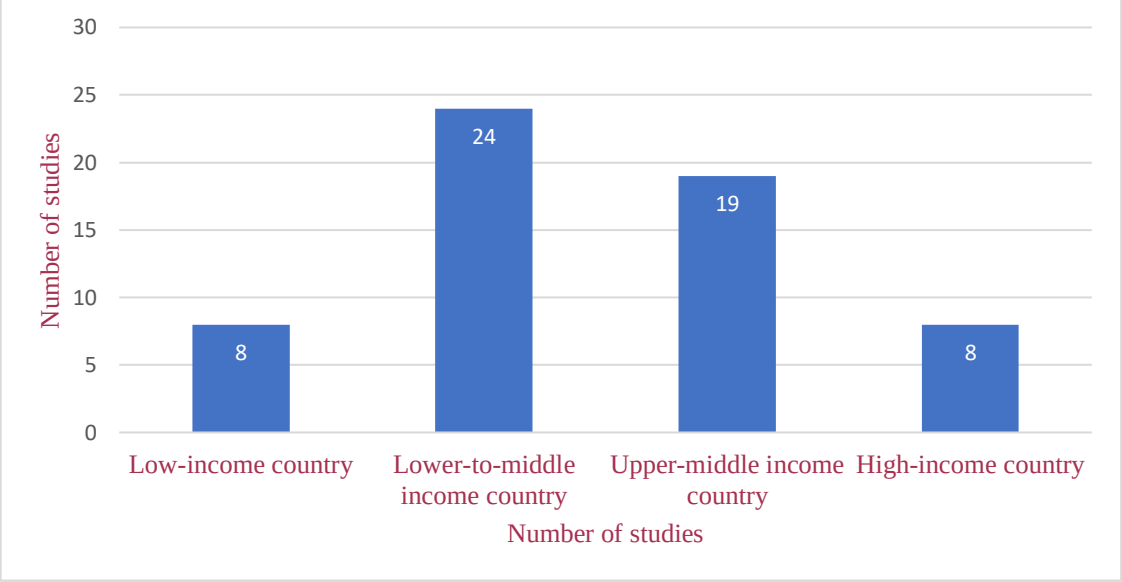
Source: Authors

Figure 5. Distribution of studies by region



Source: Authors

Figure 6. Distribution of studies by country income level



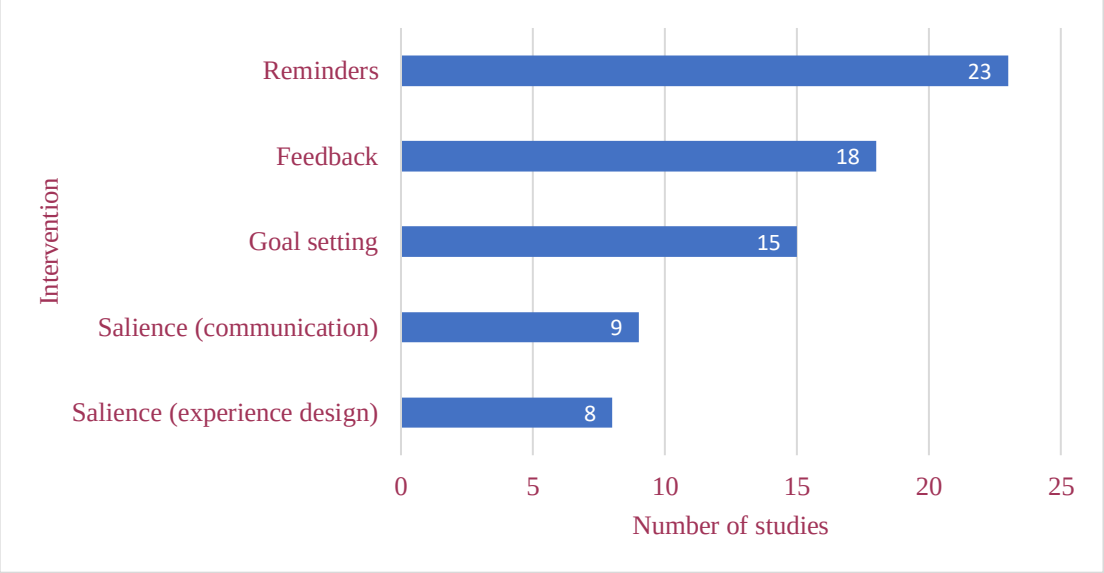
Source: Authors

c. Interventions

In this section, we describe the frequency of interventions in the evidence base displayed in Figure 7 below. While we began the extraction process with 22 intervention groups in the EGM, the scope of

the studies for the SR focuses solely on feedback, reminders, goal setting, salience (communication) and salience (experience design) with numerous multi-component interventions. Of the five interventions of interest, the most frequently reported intervention is reminders (n = 23) followed by feedback (n = 18) and goal setting (n = 15). The least reported are salience (communication) and salience (experience design), reported in nine and eight studies respectively.

Figure 7. Overview of included intervention[7] types



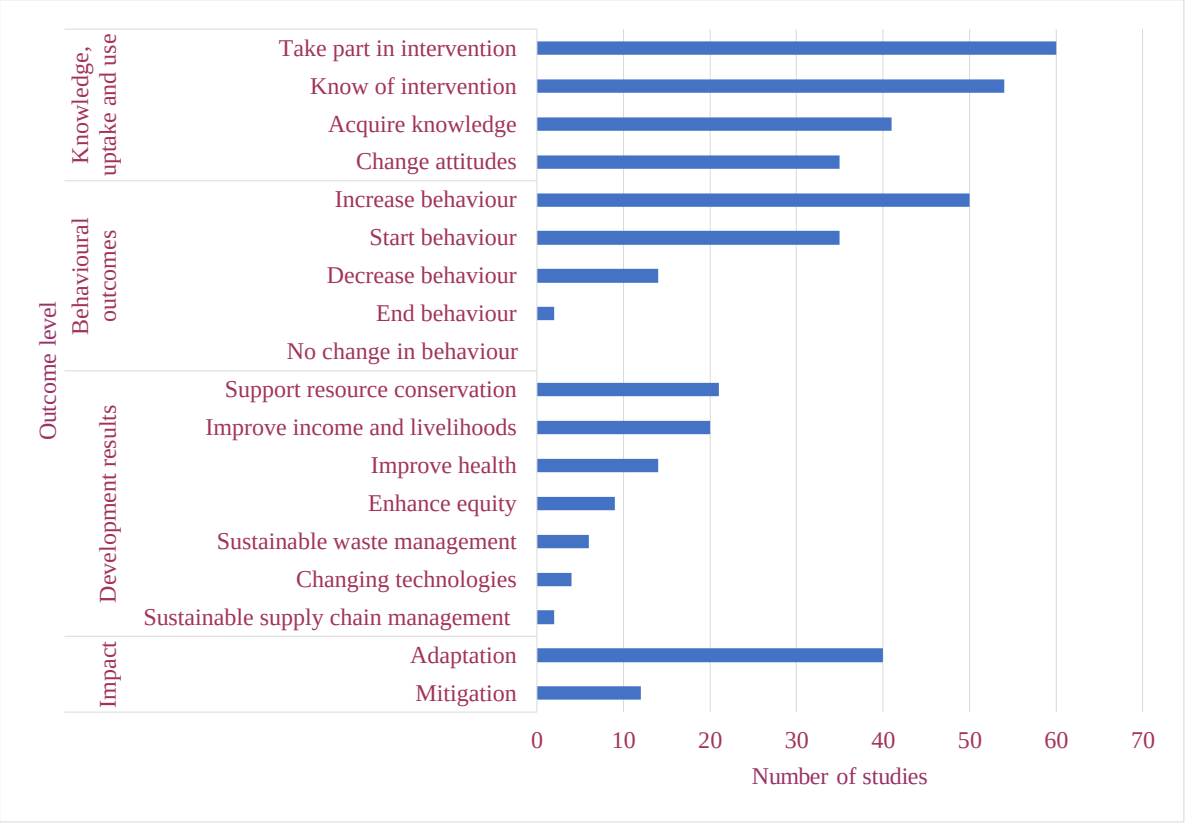
Source: Authors

d. Outcomes

Figure 8 shows the outcomes of interest reported in the included studies. Taking part in the intervention (n = 60) and knowledge of the intervention (54) are the most commonly reported intermediate outcomes followed by the acquisition of knowledge (41) and change in attitudes (35) outcomes. The second most reported outcomes fall within the behavioural outcomes level: increase in behaviour (50) and starting a behaviour (35). A limited number of studies identify with the behaviour ending outcome (2) and no study reports on the "no change in behaviour" outcome. The third most reported outcome domain is the impact[8] domain, with adaptation being reported in 40 studies, and 18 studies reporting mitigation outcomes. The development results outcome level is the least reported outcome group overall. In this outcome domain, supporting resource conservation (21) is reported the most, closely followed by improved income and livelihoods. Improved health and enhancing equity are reported in 14 and 9 studies, respectively.

[7] Studies can report multiple interventions as well as single or multicomponent interventions, leaving the cumulative total to not necessarily match the number of included studies.
[8] We classify impacts (adaptation, mitigation) using inference. For example, a behavioural outcome of reducing electricity consumption would, by inference, imply a mitigation outcome (impact).

Figure 8. Overview of included outcomes

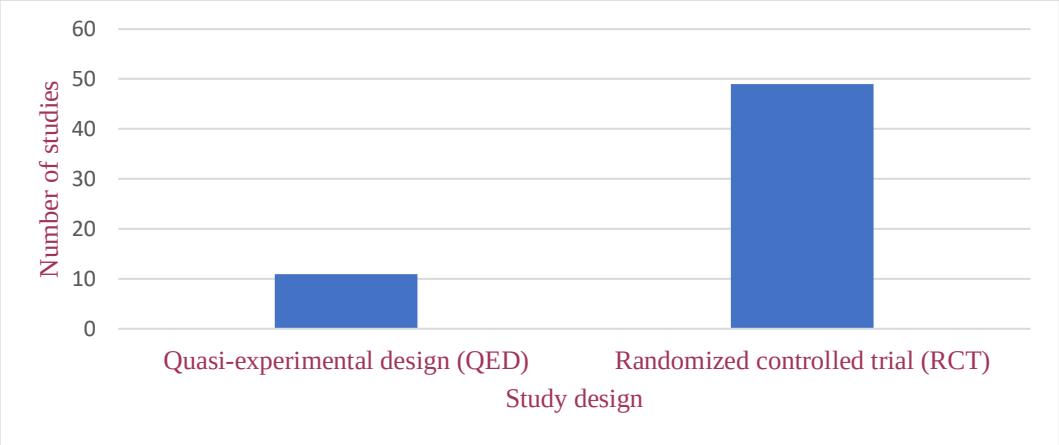


Source: Authors

e. Study design of included studies

Of the 60 included studies, a large majority of studies are RCTs (49) whereas 11 are quasi-experimental study designs (Figure 9). An element to note is that a vast majority of the included studies measure the impact of interventions during the intervention period or directly after the end of the intervention. Sixty-eight per cent (n = 41) of the included studies have a follow-up period of less than a month, 27 per cent (n = 16) have a follow-up period of 1–6 months, 3 per cent (n =2) of 7–12 months, 5 per cent (n = 3) of 19–24 months, and only 3 per cent (n = 2) over 25 months. No studies had a follow-up period of 13–18 months (n = 0).

Figure 9. Study design of included studies

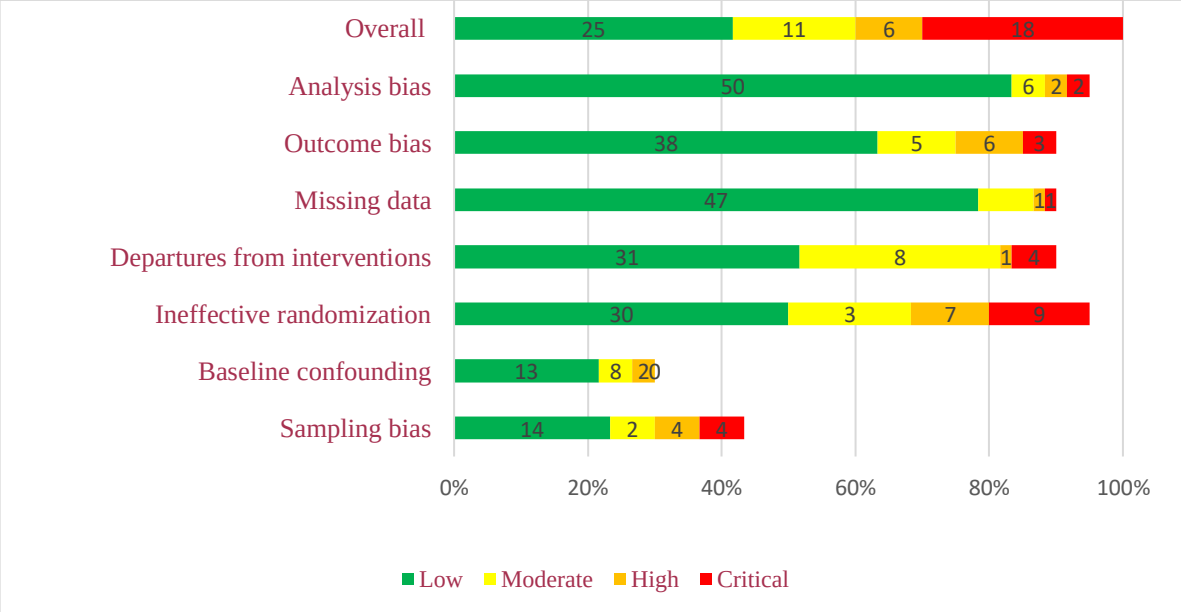


Source: Authors

B. RISK OF BIAS IN INCLUDED STUDIES

The risk of bias tool aimed to assess the risk of bias in each included study. Forty-two per cent (25/60) of the included studies were assessed to be of low risk of bias. A total of 18 studies attained a critical risk of bias rating and were excluded from the synthesis. The critical risk of bias assessment is largely due to ineffective randomization (nine studies). A total of 11 studies were rated as having a moderate risk of bias and six to have high risk, raising concerns about the reliability of the impact estimates. Throughout the synthesis, we indicate the underlying risk of bias of the studies included in the different analyses. Figure 10 provides a summary of the full risk of bias ratings of the included studies.

Figure 10. Overview of risk of bias assessment



Source: Authors

C. SYNTHESIS OF RESULTS

In this section, we report the results of our synthesis of the selected behavioural science interventions in developing countries. The synthesis is based on the results of the 42 studies assessing the impact of the five interventions included in the review. These interventions could be administered as single-component interventions and multi-component interventions either with any of the other four included interventions or other behavioural science interventions. The five interventions reported in the 42 included studies were assessed for their effectiveness on knowledge, uptake and use, behavioural outcomes, development results, and impact. A study could feature in any of these intervention categories as long as it featured an intervention in that category as a single-component or a multi-component intervention. Hence, some studies appear in more than one of these intervention categories.

We were able to conduct a statistical meta-analysis for the effects of three interventions: feedback, reminders, and goal setting – both as single interventions and in combination with other interventions. We also conducted a narrative synthesis in these interventions for studies (or treatment groups) that do not constitute the same intervention and outcome combinations. In the remaining intervention groupings, the small number of studies combined with a lack of studies with similar intervention and outcome constructs, and statistical information to calculate standardized effect sizes, prevented us from conducting a meta-analysis. We report the effects of these interventions using narrative synthesis based on a structured summary of findings tables. In total, we report 12 meta-analyses and five narrative syntheses for the above intervention groupings.

V. RESULTS OF THE META-ANALYSIS AND NARRATIVE SYNTHESIS

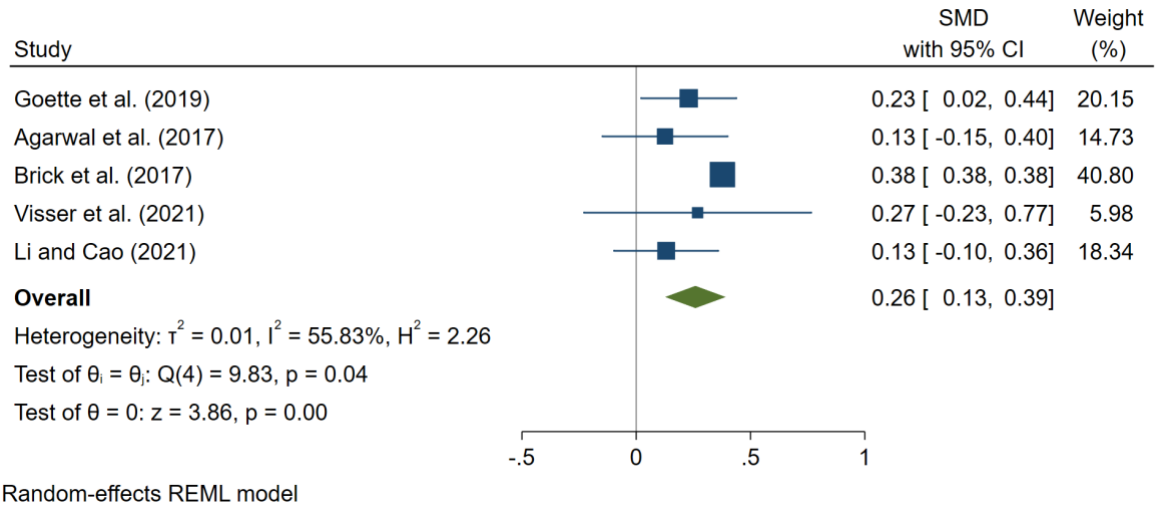
In this section, we report the results of our quantitative analysis of the effects using statistical meta-analysis and narrative synthesis. Our analysis is structured around the five intervention types and includes interventions administered as single or multi-component interventions. Using meta-analysis and narrative synthesis, we explore the effectiveness of feedback, reminders, goal setting, salience (communication), and salience (experience design).

A. INTERVENTION I: FEEDBACK

1. EFFECTS OF FEEDBACK ON BEHAVIOURAL OUTCOMES: WATER AND ELECTRICITY CONSUMPTION

We included a total of k = 5 studies, three RCTs and two quasi-experimental designs. We assessed three studies as having low risk of bias, one as having moderate risk, and one as having high risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.26 (CI: 0.13 to 0.39). The pooled effect estimate is therefore statistically significant (z=3.86, p<0.00). This implies that feedback interventions have a significant positive effect on water and electricity consumption. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 11 below. According to the Q-test, the meta-analysis results are subject to a moderate degree of heterogeneity (Q (4) =9.83, p=0.04, $\hat{\tau}^2$=0.01, I^2=55.83 per cent), and the CIs of all but one study overlap. To assess the robustness of the identified effect, we next report the results of our sensitivity and moderator analyses, which investigate whether the observed overall effect might be driven by variables other than the applied reminder interventions. Applying the GRADE framework, the evidence on the effects of feedback on behavioural outcomes (water and electricity consumption) is of moderate quality (see Appendix 4). With five studies, tests of publication bias are not valid, as discussed in section IIIG.

Figure 11. Meta-analysis of feedback on behavioural outcomes: water and electricity consumption



Source: Authors

Sensitivity analysis of feedback on behavioural outcomes: water and electricity

We investigated whether the variance in effect sizes might be caused by factors related to the applied evaluation design (i.e. study type, risk of bias and period of follow-up). For instance, a more rigorous evaluation approach might yield systematically different effect sizes from those with a less robust evaluation design. We, therefore, investigated the sensitivity of our pooled effect estimate to the above design factors. Table 1 presents an overview of how the meta-analysis results vary if different groups of studies are combined according to the above design features. Differences in the pooled effect size for each variable could indicate that the overall results of the meta-analysis are sensitive to the design variable under investigation. It is, however, important to note that Table 1 presents merely an observational approach to uncover possible sensitivities that we then formally assessed statistically using a one-way random effects restricted maximum likelihood (REML) model.

Table 1. Sensitivity analysis of feedback on behavioural outcomes: water and electricity

Variable	SMD	95% CI	Q	Tau^2	I^2	p-value (Q)	Sample	Sensitivity
Feedback: all studies	0.26	0.13 to 0.39	9.83	0.01	55.85%	0.04	5	
Design			4.13			0.04		Sensitive
RCT	0.34	0.23 to 0.45	2.11	0.00	29.86%	0.35	3	
Quasi-experimental design	0.13	-0.05 to 0.30	0.00	0.00	0.00%	0.98	2	
Risk of bias			0.94			0.63		Not sensitive
Low risk of bias	0.28	0.13 to 0.43	6.42	0.01	66.37%	0.04	3	
Moderate risk of bias	0.27	-0.23 to 0.77	0.00	0.00	0.00%	0.00	1	
High risk of bias	0.13	-0.15 to 0.40	0.00	0.00	0.00%	0.00	1	
Period of follow-up	Lack of sufficient evidence in each category for sensitivity analysis on this variable.[9]							N/A
Less than a month	0.26	0.13 to 0.39	9.83	0.01	55.85%	0.04	5	
1–6 months	N/A	No observations					0	
7–12 months	N/A	No observations					0	
13–18 months	N/A	No observations					0	
19–24	N/A	No					0	

[9] All studies have a follow-up period of less than a month. One study measured outcomes immediately at the end of the intervention period and four studies assessed changes in outcome during the intervention period.

Variable	SMD	95% CI	Q	Tau^2	I^2	p-value (Q)	Sample	Sensitivity
months		observations						
>25 months	N/A	No observations					0	

Source: Authors

In our meta-analysis on feedback on water and electricity consumption, we combined results from RCTs and quasi-experimental designs. The results of the meta-analysis were sensitive to the applied study design (Q=4.13; p=0.04), with RCTs yielding a higher pooled effect estimate. While we observed that a lower risk of bias led to larger effects, testing for the significance of this difference in effect sizes established that variances in the quality of studies did not influence the overall results of the meta-analysis (Q= 0.94; p=0.63). We could not test whether differences in the period of follow-up systematically affected the pooled effect due to a lack of sufficient studies in each category for this test (Q=1.44; p=0.23).

Moderator analysis of feedback on behavioural outcomes: water and electricity consumption

In addition to assessing whether variables relating to study design influenced the robustness of the meta-analysis, we further conducted moderator analyses on a combination of extrinsic characteristics, and substantive characteristics listed in Table 2. A context characteristic, such as income may influence the intervention effects. For example, programmes in upper-middle income countries might be more effective than those in low-income countries, and so forth. We had sufficient data to test most moderators for this intervention type and investigated variables related to publication type and date, and context characteristics of the intervention that might systematically moderate intervention effects identified in the meta-analysis (Table 2). We lacked sufficient data to test for the influence of intervention length.

When reporting the moderator analysis, we used the same structure as in the sensitivity analysis based on an observational overview table followed by a one-way REML model. Table 2 presents an overview of how the meta-analysis results vary if different groups of studies are combined according to the moderator variables under investigation. Differences in the pooled effect size for each variable could indicate that the overall results of the meta-analysis are sensitive to the moderator variable under investigation, which is then formally tested in the REML model.

Publication type and date, income groups, and whether the programme participants received incentives did not significantly influence the results of our meta-analysis. However, the location of the programme and the implementing agency has influenced the results of our meta-analysis. Country-wise, we observe larger effects in South Africa, followed by Singapore and lastly, China. Region-wise, the effects are larger in Sub-Saharan Africa relative to East Asia and the Pacific. We observe the largest effects for programmes that were implemented by government agencies followed by interventions implemented by academic institutions and the least amount of effects in programmes that have no specific implementing agency.

Table 2. Moderator analysis of feedback on behavioural outcomes: water and electricity consumption

Variable	SMD	95% CI	Q	Tau^2	I^2	p-value (Q)	Sample	Sensitivity
Feedback: all	0.26	0.13 to	9.83	0.01	55.85	0.04	5	

Variable	SMD	95% CI	Q	Tau²	I²	p-value (Q)	Sample	Sensitivity
studies		0.39			%			
Publication type			1.01			0.31		Not sensitive
Academic journal article	0.28	0.15 to 0.42	6.60	0.01	54.85 %	0.09	4	
Research report	0.13	-0.15 to 0.40	0.00	0.00	0.00%	0.00	1	
Publication date			0.74			0.69		Not sensitive
2017	0.29	0.06 to 0.52	3.24	0.02	69.10 %	0.07	2	
2019	0.23	0.02 to 0.44	0.00	0.00	0.00%	0.00	1	
2021	0.15	-0.05 to 0.36	0.24	0.00	0.00%	0.62	2	
Length of intervention	Lack of sufficient evidence in each category for moderator analysis							N/A
Less than a month	N/A	No observations					0	
1–6 months	0.26	0.13 to 0.39	9.83	0.01	55.85 %	0.04	5	
7–12 months	N/A	No observations					0	
13–18 months	N/A	No observations					0	
19–24 months	N/A	No observations					0	
>25 months	N/A	No observations					0	
Country			9.30			0.04		Sensitive
China	0.13	-0.10 to 0.36	0.00	.	0.00%	.	1	
Singapore	0.19	0.03 to 0.36	0.35	0.00	0.00%	0.55	2	
South Africa	0.38	0.37 to 0.38	0.18	0.00	0.00%	0.67	2	
World Bank region			9.12			0.00		Sensitive
Middle East and North Africa	N/A	No observations						
East Asia and Pacific	0.17	0.04 to 0.30	0.53	0.00	0.00%	0.77	3	
Sub-Saharan Africa	0.38	0.37 to 0.38	0.18	0.00	0.00%	0.67	2	
Europe and Central Asia	N/A	No observations						
Latin America and Caribbean	N/A	No observations						

VARIABLE	SMD	95% CI	Q	TAU^2	I^2	P-VALUE (Q)	SAMPLE	SENSITIVITY
South Asia	N/A	No observations						
Not Specified	N/A	No observations						
Income group			0.58			0.44		Not sensitive
Low-income country (s)	N/A	No observations						
Low-middle-Income (s)	N/A	No observations						
Upper-middle income country (s)	0.29	0.10 to 0.47	4.68	0.02	58.18 %	0.10	3	
High income country (s)	0.19	0.03 to 0.36	0.35	0.00	0.00%	0.55	2	
Incentives			0.58			0.44		Not sensitive
No	0.29	0.10 to 0.47	4.68	0.02	58.18 %	0.10	3	
Yes	0.19	0.03 to 0.36	0.	0.00	0.00%	0.55	2	
Implementing agency			9.81			0.01		Sensitive
Academic institution	0.24	0.004 to 0.43	0.02	0.00	0.00%	0.89	2	
Government agency	0.38	0.38 to 0.38	0.00	0.00	0.00%	0.00	1	
Not specified	0.13	-0.05 to 0.30	0.00	0.00	0.00%	0.98	2	
Charitable or private foundation	N/A	No observations						
For-profit firm	N/A	No observations						
International aid agency	N/A	No observations						
International financial institution	N/A	No observations						
Non-profit organization	N/A	No observations						
Think tank	N/A	No observations						

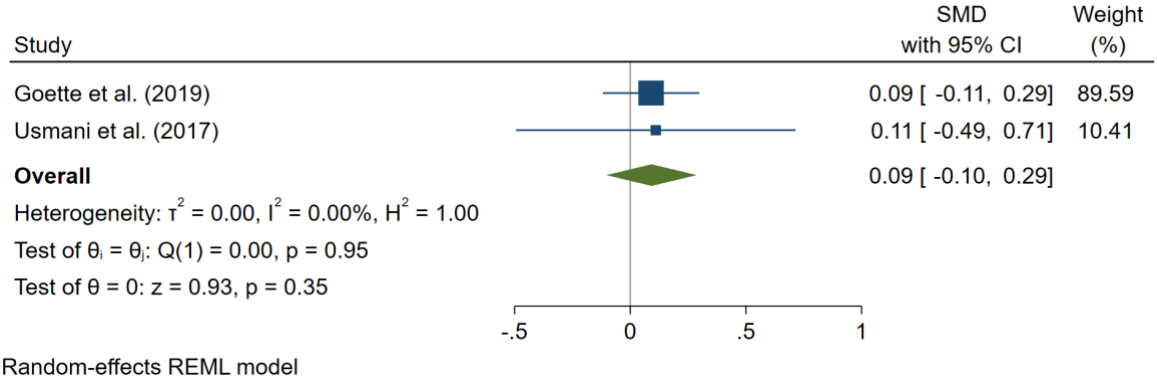
Source: Authors

2. EFFECTS OF FEEDBACK AND MICRO-INCENTIVES ON BEHAVIOURAL OUTCOMES: WATER AND ELECTRICITY CONSUMPTION

Only two studies (RCTs) reported the impact of feedback and micro-incentives on similar behavioural outcomes, with one study reporting water consumption and the other electricity consumption – thus we included $k = 2$ studies in the analysis. We assessed one of the studies as having low risk of bias and the other as having moderate risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.09 (CI: -0.10 to 0.29). The pooled effect estimate is therefore statistically insignificant (z=0.93, p<0.35). Interventions with combined feedback and micro-incentives present a small positive but statistically insignificant effect on water and electricity consumption. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 12.

Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was no significant amount of heterogeneity in the true outcomes (Q (1) =0.93, p=0.95, $\hat{\tau}^2$=0.00, I^2=0.00 per cent). With only two studies, and given there was no heterogeneity among the effects, sensitivity moderator analyses were not possible, and tests of publication bias are not valid. The GRADE quality of evidence assessment indicates that evidence on the effects of feedback and micro-incentives on behavioural outcomes (water and electricity consumption) is of very low quality (see Appendix 4).

Figure 12. Meta-analysis of feedback and micro-incentives on behavioural outcomes: water consumption and solid-fuel use



Source: Authors

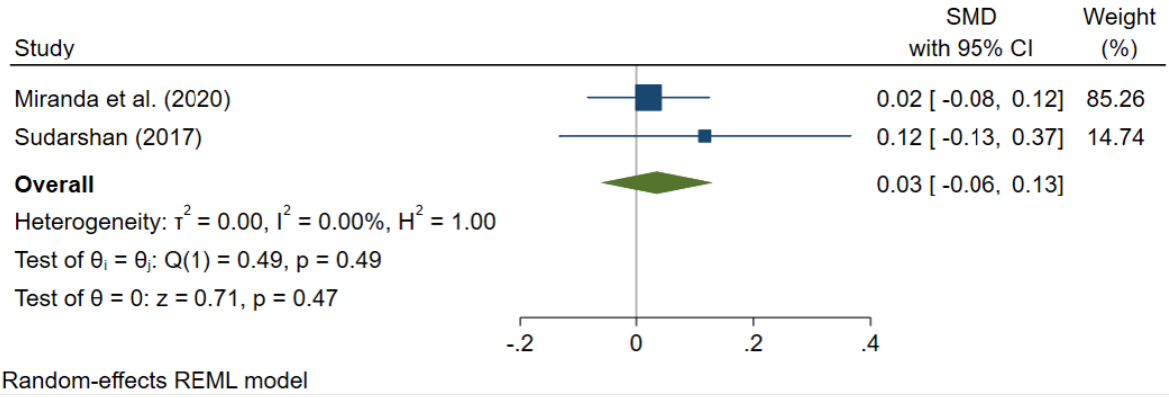
3. EFFECTS OF FEEDBACK AND SOCIAL BENCHMARKING ON BEHAVIOURAL OUTCOMES: WATER AND ELECTRICITY CONSUMPTION

We included $k = 2$ studies in the analysis of the combined effect of feedback and social benchmarking on behavioural outcomes in water and electricity consumption, both RCTs. We assessed both studies as having low risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.03 (CI: -0.06 to 0.13). The pooled effect estimate is therefore statistically insignificant (z=0.71, p=0.47), indicating that interventions combining feedback and social benchmarking present a small positive but statistically insignificant effect on water and electricity

consumption outcomes. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 13.

Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was no significant amount of heterogeneity in the true outcomes (Q (1) =0.49, p=0.49, $\hat{\tau}^2$=0.00, I^2=0.00 per cent). With only two studies, and given there was no heterogeneity among the effects, sensitivity and moderator analyses were not possible, and tests of publication bias are not valid. Applying GRADE indicates that the evidence on the effects of feedback and social benchmarking on behavioural outcomes (water and electricity consumption) is of very low quality (see Appendix 4).

Figure 13. Meta-analysis of feedback and social benchmarking on behavioural outcomes: water consumption and electricity consumption

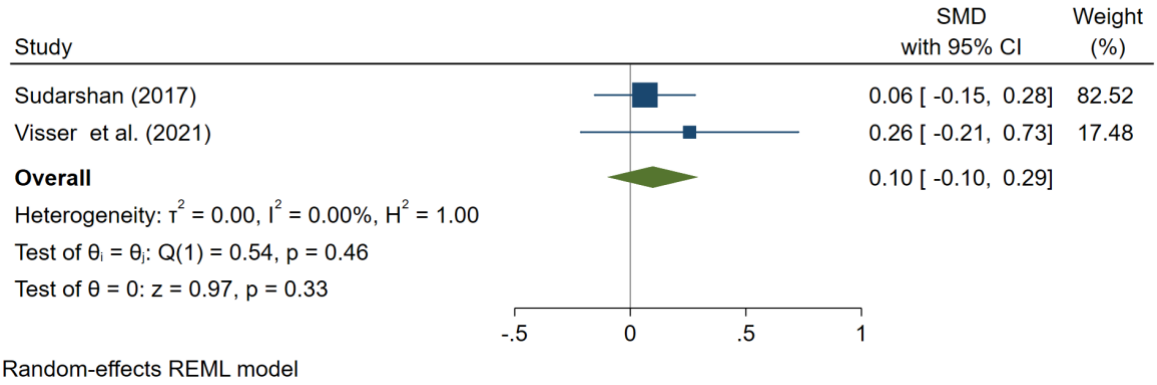


Source: Authors

4. EFFECTS OF FEEDBACK, SOCIAL BENCHMARKING AND MICRO-INCENTIVES ON BEHAVIOURAL OUTCOMES: WATER AND ELECTRICITY CONSUMPTION

Two studies reported the impact of combined feedback, social benchmarking and micro-incentives on behavioural outcomes (k = 2 studies), both RCTs. One study reported water consumption and the other electricity consumption as an outcome. We assessed one of the studies as having low risk of bias and the other as having moderate risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.10 (CI: -0.10 to 0.29) and is therefore statistically insignificant (z=0.97, p<0.33), indicating that combining these three interventions presents a small positive but statistically insignificant effect on water and electricity consumption. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 14. Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was no significant amount of heterogeneity in the true outcomes (Q (1) =0.54, p=0.46, $\hat{\tau}^2$=0.00, I^2=0.00 per cent). Given there were only two studies, and that there was no heterogeneity among the effects, sensitivity and moderator analyses were not possible, and tests of publication bias are not valid. The GRADE quality of evidence assessment indicates that evidence on the effects of feedback, social benchmarking and micro-incentives on behavioural outcomes (water and electricity consumption) is of very low quality (see Appendix 4).

Figure 14. Meta-analysis of feedback, social benchmarking and micro-incentives on behavioural outcomes: water and electricity consumption



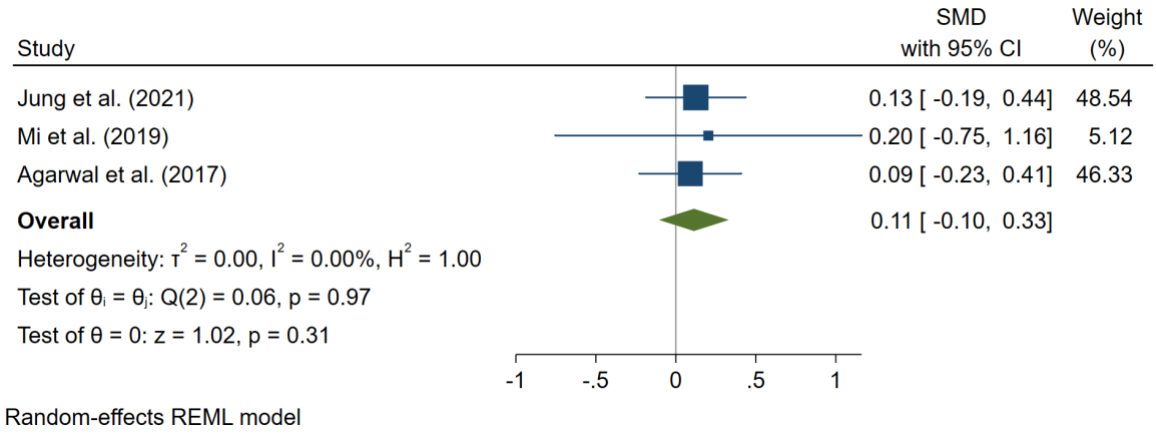
Source: Authors

5. EFFECTS OF FEEDBACK AND GOAL SETTING ON BEHAVIOURAL OUTCOMES: ELECTRICITY CONSUMPTION

We included k = 3 studies in the analysis of the combined effect of feedback and goal-setting on behavioural outcomes in electricity consumption. All studies are RCTs. Two studies were assessed as having low risk of bias whilst one had a high risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.11 (CI: −0.10 to 0.43), and is therefore statistically insignificant (z=1.02, p=0.31). Combined feedback and goal-setting interventions present a small positive but statistically insignificant effect on electricity consumption behaviours. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 15.

Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was no significant amount of heterogeneity in the true outcomes (Q (1) =0.06, p=0.97, $\hat{\tau}^2$=0.00, I^2=0.00 per cent). With only two studies, and given there was no heterogeneity among the effects, moderator analyses were not possible, and tests of publication bias are not valid. The GRADE quality of evidence assessment shows that evidence on the effects of feedback and goal-setting interventions on behavioural outcomes (for electricity consumption) is of very low quality (see Appendix 4).

Figure 15. Meta-analysis of feedback and goal setting on behavioural outcomes: electricity consumption



Source: Authors

Sensitivity analysis of feedback and goal setting on electricity consumption

We investigated whether the variance in effect sizes might be caused by factors related to the applied evaluation design (i.e. study type, risk of bias and period of follow-up). Table 3 presents an overview of how the meta-analysis results vary if different groups of studies are combined according to the above design features. The results of the meta-analysis are not significantly influenced by the risk of bias of the included studies. We did not have sufficient information to test for the sensitivity of our meta-analysis results to study design and follow-up period, as all three studies fall into one category of each of the variables.

Table 3. Sensitivity analysis of feedback and goal setting on electricity consumption

VARIABLE	SMD	95% CI	Q	TAU²	I²	P-VALUE (Q)	SAMPLE	SENSITIVITY
Feedback: all studies	0.11	-0.10 to 0.33	0.06	0.00%	0.00%	0.97	3	
Design	Lack of sufficient evidence in each category for sensitivity analysis							N/A
RCT	0.11	-0.10 to 0.33	0.06	0.00%	0.00%	0.97	3	
Quasi-experimental design	N/A	No observations					0	
Risk of bias			0.04		0.85			Not sensitive
Low risk of bias	0.13	-0.16 to 0.43	0.02	0.00%	0.00	0.88	2	
Moderate risk of bias	N/A	No observations					0	
High risk of bias	0.09	-0.23 to 0.41	0.00	0.00%	0.00%	0.00	1	
Period of follow-up	Lack of sufficient evidence in each category for sensitivity analysis							N/A

VARIABLE	SMD	95% CI	Q	TAU^2	I^2	P-VALUE (Q)	SAMPLE	SENSITIVITY
Less than a month	0.11	-0.10 to 0.33	0.06	0.00%	0.00%	0.97	3	
1–6 months	N/A	No observations					0	
7–12 months							0	
13–18 months	N/A	No observations					0	
19–24 months	N/A	No observations					0	
>25 months	N/A	No observations					0	

Source: Authors

6. NARRATIVE SYNTHESIS OF FEEDBACK ON BEHAVIOURAL OUTCOMES

We also investigated whether a narrative synthesis of feedback interventions is feasible for interventions and outcome combinations not covered in the meta-analysis. However, due to intervention heterogeneity in the intervention categories these studies could not be synthesized, as can be seen in Table 4. A narrative overview of the individual studies and key findings on impact are provided in Table 4, and more details on the interventions and main findings on the effects are highlighted in Appendix 5.

Table 4. Narrative synthesis of feedback on behavioural outcomes interventions

STUDY	INTERVENTION TREATMENT COMPONENTS	CONTEXT	FINDINGS
Goette and others (2019) Low risk of bias	Feedback and lottery	East Asia and the Pacific: Singapore Income group: High income country	Higher baseline water consumption households have a much higher conservation effect from the treatment. Households with a 1-litre higher consumption in baseline use, reduced water use by around 0.066 litres per household (SMD: 0.0961, CI: -0.1089 to 0.3011). Lower baseline households seem to increase their water use after treatment.
Barido and others (2018) Moderate risk of bias	Feedback, reminders and micro-incentives	Latin America and the Caribbean: Nicaragua Income group: Low-income country	Houses and micro-enterprises assigned to the intervention reduced their energy consumption relative to the control group during the intervention period months, as compared to the previous year (SMD: -1.2252, CI: -1.7767 to -0.6737).
Brick and others (2017) Low risk of bias	Feedback and framing devices	Sub-Saharan Africa: South Africa	The financial gain messages treatment reduced water consumption by 206 litres per household per month on average (SMD: 0.0094, CI: -0.004 to 0.0228), and

Study	Intervention treatment components	Context	Findings
		Income group: Upper-middle-income country	in the financial loss messages treatment reduced water consumption by 181 litres per household per month on average (SMD: -0.0076, CI: -0.4173 to 0.4326).
Ruiz-Tagle and Schueftan (2021) Low risk of bias	Feedback	Latin America and the Caribbean: Chile Income group: High-income country	**Emissions of fine particulate matter (PM2.5)** Emissions of fine particulate matter (PM2.5): average emissions (imputed grams/hour) for those in the treatment group declined by 10.8% after installation of the treatment (SMD: -0.5117, CI: -1.171 to 0.1476) **A decline in imputed wood stove emissions,** by a coefficient of 1.844 (SMD: 0.3824, CI: 0.2727 to 1.0376) The **probability of shifting damper towards less polluting settings** increased by 0.603 (SMD:1.6168, CI: 0.8768 to 2.3569)
Thondhlana (2016) Moderate risk of bias	Feedback, reminders and rules of thumb	Sub-Saharan Africa: South Africa Income group: Upper middle income	There was a significant reduction (24.5 KWh) in electricity consumption during the intervention period in the treatment groups. However, in the control group there was no significant difference (4.81 KWh) in electricity consumption between the same time points.
Torres and Carlsson (2016) High risk of bias	Feedback, social norms and default	Latin America and the Caribbean: Colombia Income group: Upper-middle-income country.	The average household participating in the experiment reduced water use by 13% (SMD: 0.1152, CI: - 0.0012 to 0.2317) and 6.3% (SMD: - 0.0505, CI: -0.0658 to 0.1669) in the first 6 and 11 months after the start of the experiment, respectively.

Source: Authors

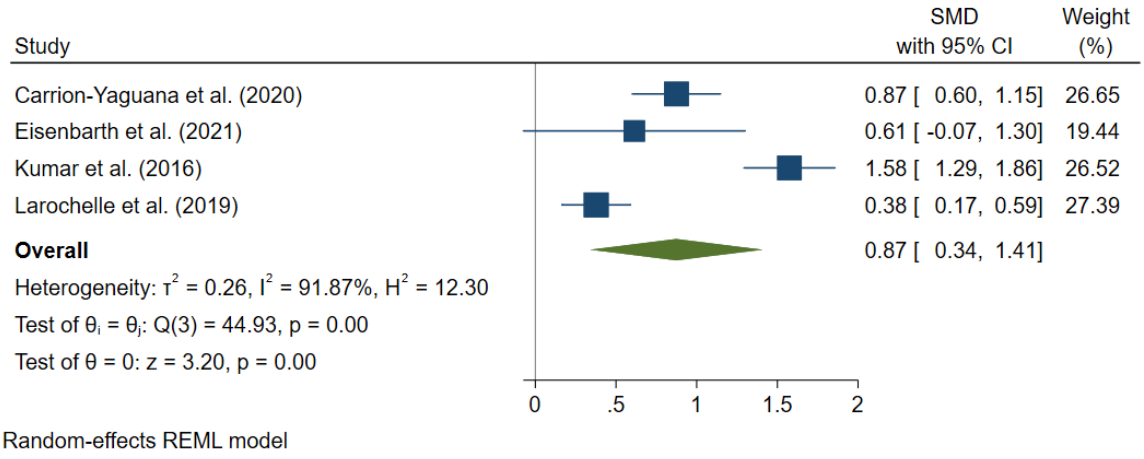
B. Intervention II: Reminders

1. Effects of reminders on knowledge, uptake and use: acquisition of knowledge

We identified and included a total of k = 4 studies, three RCTs and one quasi-experimental study, that assessed the impact of reminders on knowledge acquisition. Two studies were assessed as having a low risk of bias and two as having high risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.87 (CI: 0.34 to 1.41), and is therefore statistically significant and differed significantly from zero (z=3.20, p<0.00). This indicates that reminder interventions resulted in a large positive and statistically significant effect on the acquisition of knowledge. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 16 below.

According to the Q-test, there is a significant amount of heterogeneity in the true outcomes (Q (1) =44.93, p=0.00, $\hat{\tau}^2$=0.26, I^2=91.87 per cent), and the CIs of the two studies do not overlap. To assess the robustness of the identified effect, we next report the results of our sensitivity and moderator analyses, which investigate whether the observed overall effect might be driven by variables other than the applied reminder interventions. With four studies, tests of publication bias are not valid. Applying the GRADE framework, the evidence on the effects of reminders on knowledge, uptake and use outcomes (knowledge acquisition) is of very low quality (see Appendix 4).

Figure 16. Meta-analysis of reminders on knowledge, uptake and use: acquisition of knowledge



Source: Authors

Sensitivity analysis of reminders on knowledge, uptake and use: acquisition of knowledge

We investigated whether the variance in effect sizes might be caused by factors related to the applied evaluation design (i.e. study type, risk of bias and period of follow-up). Table 5 presents an overview of how the meta-analysis results vary if different groups of studies are combined according to the above design features. Differences in the pooled effect size for each variable could indicate that the overall results of the meta-analysis are sensitive to the design variable under investigation. In our meta-analysis on reminders for the acquisition of knowledge, we combined results from three RCTs and one quasi-experimental design. The results of the meta-analysis were sensitive to the applied study design (Q=17.15; p=0.00) and we observed larger effects from quasi-experimental designs relative to the RCTs. Regarding pooling studies with different risks of bias, we observed that a higher risk of bias led to larger effects, but testing for the significance of this difference in effect sizes established that variances in the quality of studies did not influence the overall results of the meta-analysis (Q=0.96; p=0.33). We further tested whether differences in the period of follow-up systematically affected the pooled effect size and found that the results are sensitive to follow-up. The effects are higher for studies with a follow-up of 1–6 months compared to a follow-up period of less than a month (Q=17.15; p=0.00).

Table 5. Sensitivity analysis of reminders on knowledge, uptake and use: acquisition of knowledge

VARIABLE	SMD	95% CI	Q	TAU^2	I^2	P-VALUE (Q)	SAMPLE	SENSITIVITY
Reminders: all studies	0.87	0.34 to 1.41	44.93	0.26	91.87%	0.00	4	
Design			17.15			0.00		Not sensitive
RCT	0.61	0.26 to 0.97	8.00	0.07	71.36%	0.02	3	
Quasi-experimental design	1.58	1.29 to 1.86	0.00	0.00	0.00	0.00	1	
Risk of bias			0.96			0.33		Not sensitive
Low risk of bias	0.62	0.13 to 1.11	7.98	0.11	87.48%	0.00	2	
Moderate risk of bias	N/A	No observations						
High risk of bias	1.15	0.21 to 2.08	6.44	0.39	84.46%	0.01	2	
Period of follow-up			17.15			0.00		Sensitive
Less than a month	0.61	0.26 to 0.97	8.00	0.07	71.36%	0.02	3	
1–6 months	1.58	1.29 to 1.86	00.00	00.00	0.00%	0.00	1	
7–12 months	N/A	No observations						
13–18 months	N/A	No observations						
19–24 months	N/A	No observations						
>25 months	N/A	No observations						

Source: Authors

Moderator analysis of reminders on knowledge, uptake and use: acquisition of knowledge

We then assessed whether a range of moderators influenced the robustness of the meta-analysis. We had sufficient data to test most moderators for this intervention and outcome group, and investigated variables related to publication type and date, intervention length, and context characteristics of the intervention that might systematically moderate intervention effects identified in the meta-analysis (see Table 6). We find that context (country, region and income group) is a significant moderator. Specifically, amongst the three countries in which the evaluations were conducted, the largest effects were realized in Pakistan as compared to Ecuador and Uganda. Regionally, this translates to South Asia witnessing the largest effects. Correspondingly, the same pattern is noticed for income group, with the largest effects being realized in a low-middle-income country. We also find that providing incentives for participants in the intervention target group is a significant moderator. The

effects are larger for evaluations where participants are not incentivized to participate than for those where incentives for participation are provided. Length of intervention and implementing agency type are not significant moderators. There is a lack of sufficient data to conduct a moderator analysis on publication type and date.

Table 6. Moderator analysis of reminders on knowledge, uptake and use: acquisition of knowledge

VARIABLE	SMD	95% CI	Q	TAU^2	I^2	P-VALUE (Q)	SAMPLE	SENSITIVITY
Reminders: all studies	0.87	0.34 to 1.41	44.93	0.26	91.87%	0.00	4	
Publication type	Lack of sufficient evidence in each category for sensitivity analysis							N/A
Academic journal article	N/A						4	
Research report	N/A	No observations					0	
Publication date	Lack of sufficient evidence in each category for sensitivity analysis							N/A
2016	N/A						1	
2019	N/A						1	
2020	N/A						1	
2021	N/A						1	
Length of intervention		0.43				0.51		Not sensitive
Less than a month	N/A	No observations						
1–6 months	0.94	0.26 to 1.62	44.57	0.35	95.38%	0.00	3	
7–12 months	0.61	-0.07 to 1.30	0.00	0.00	.	.	1	
13–18 months	N/A	No observations						
19–24 months	N/A	No observations						
>25 months	N/A	No observations						
Country			14.87			0.00		Sensitive
Ecuador	0.62	0.13 to 1.11	7.98	0.11	87.48%	0.00	2	
Pakistan	1.58	1.29 to 1.86	0.00	0.00	0.00%	0.00	1	
Uganda	0.61	-0.07 to 1.30	0.00	0.00	0.00%	0.00	1	
World Bank region			14.87			0.00		Sensitive
Latin America and Caribbean	0.62	0.13 to 1.11	7.98	0.11	87.48%	0.00	2	
Sub-Saharan Africa	0.61	-0.07 to 1.30	0.00	0.00	0.00%	0.00	1	

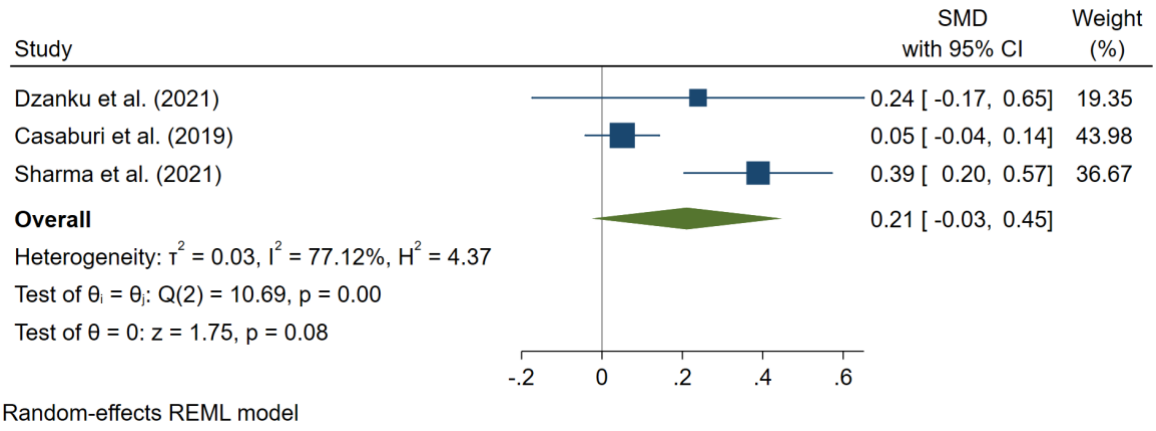
VARIABLE	SMD	95% CI	Q	TAU^2	I^2	P-VALUE (Q)	SAMPLE	SENSITIVITY
South Asia	1.58	1.29 to 1.86	0.00	0.00	0.00%	0.00	1	
East Asia and Pacific	N/A	No observations						
Middle East and North Africa	N/A	No observations						
Europe and Central Asia	N/A	No observations						
Not Specified	N/A	No observations						
Income group			14.87			0.00		Sensitive
Low-income country (s)	0.61	-0.07 to 1.30	0.00	0.00	0.00%	0.00	1	
Low-middle-income country (s)	1.58	1.29 to 1.86	0.00	0.00	0.00%	0.00	1	
Upper middle-income country (s)	0.62	0.13 to 1.11	7.89	0.11	87.48%	0.00	2	
High-income country (s)	N/A	No observations						
Incentives			5.09			0.02		Sensitive
No	1.07	0.51 to 1.63	15.00	0.20	86.04%	0.00	3	
Yes	0.38	0.17 to 0.59	0.00	0.00	0.00%	0.00	1	
Implementing agency			0.00			0.99		Not sensitive
Charitable or private foundation	0.87	0.60 to 1.15	0.00	0.00	0.00%	0.00	1	
Not Specified	0.87	0.11 to 1.63	44.72	0.40	93.69%	0.00	3	
Academic institution	N/A	No observations						
For-profit firm	N/A	No observations						
Government agency	N/A	No observations						
International aid agency	N/A	No observations						
International financial institution	N/A	No observations						
Non-profit organisation	N/A	No observations						
Think tank	N/A	No observations						

Source: Authors

2. EFFECTS OF REMINDERS ON DEVELOPMENT RESULTS: CROP YIELDS

Three studies reported the impact of reminders on development results in crop yields (k = 3 studies). All are RCTs. We assessed two of the studies as having low risk of bias and the other as having moderate risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.21 (CI: -0.03 to 0.45). Therefore, the pooled effect estimate is statistically not significant (z=1.75, p<0.08), implying that reminders do have a positive but statistically insignificant effect on farmer crop yields. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 17. Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was a significant amount of heterogeneity in the true outcomes (Q (1) =10.69, p=0.00, $\hat{\tau}^2$=0.03, I^2=77.12 per cent). Given that the meta-analysis only includes three studies, moderator analyses were not possible, but we conducted a sensitivity analysis to test if the variance in effect sizes might be caused by factors related to the applied evaluation design (i.e. study type, risk of bias and period of follow-up). The GRADE framework rated the evidence on the effects of reminders on development results in crop yields as being of very low quality (see Appendix 4).

Figure 17. Meta-analysis of reminders on development results: crop yields



Source: Authors

Sensitivity analysis of reminders on development results: crop yields

There are insufficient studies in the design categories to test the systematic influence of study design, as all the studies are RCTs. However, the results are systematically influenced by the risk of bias of the studies, and the period of follow-up. Larger effects are noticed in the studies with a moderate risk of bias relative to studies with low risk of bias, and larger effects are noticed for studies with a follow-up of less than a month compared to studies with a 1–6 month follow-up interval. Table 7 presents the results of the sensitivity analysis for reminders on development results in crop yields.

Table 7. Sensitivity analysis of reminders on development results: crop yields

VARIABLE	SMD	95% CI	Q	TAU^2	I^2	P-VALUE (Q)	SAMPLE	SENSITIVITY
Reminders: all studies	0.21	-0.03 to 0.45	10.69	0.03	77.12%	0.00	3	
Design	Lack of sufficient evidence in each category for sensitivity analysis							N/A
RCT	0.21	-0.03 to 0.45	10.69	0.03	77.12%	0.00	3	
Quasi-experimental design	N/A	No observations					0	
Risk of bias			9.93			0.00		Sensitive
Low risk of bias	0.06	-0.03 to 0.15	0.76	0.00	0.00%	0.38	2	
Moderate risk of bias	0.39	0.20 to 0.57	0.00	0.00	0.00%	0.00	1	
High risk of bias	N/A	No observations						
Period of follow-up			9.93			0.00		Sensitive
Less than a month	0.39	0.20 to 0.57	0.00	0.00%	0.00%	0.00	1	
1–6 months	0.06	-0.03 to 0.15	0.76	0.00%	0.00	0.38	2	
7–12 months	N/A	No observations						
13–18 months	N/A	No observations						
19–24 months	N/A	No observations						
>25 months	N/A	No observations						

Source: Authors

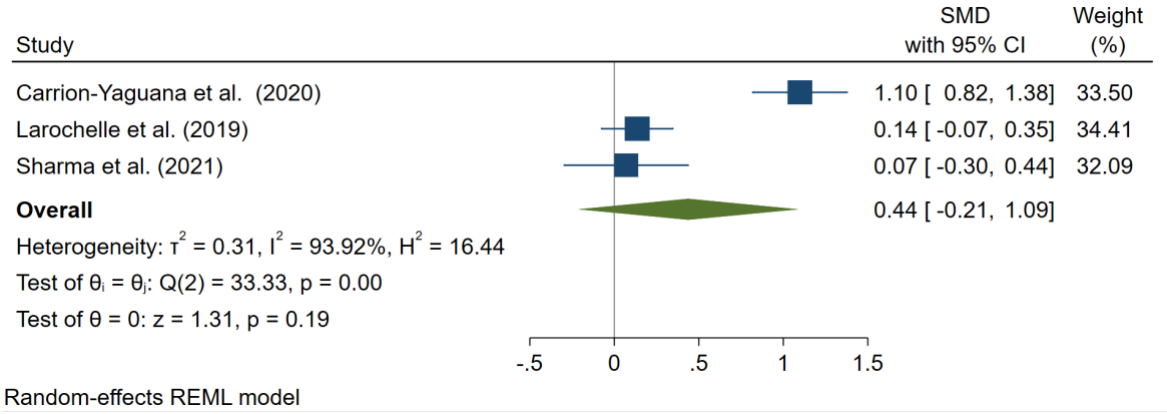
3. EFFECTS OF REMINDERS ON DEVELOPMENT RESULTS: IMPROVED INCOME AND LIVELIHOODS

We included three studies in the meta-analysis of reminders on development results (k = 3), and all studies present RCTs. We assessed two of the studies as having low risk of bias and the other as having moderate risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.44 (CI: -0.21 to 1.09), and therefore is statistically insignificant (z=1.31, p<0.19) – that is, reminders present a positive but statistically insignificant effect on development results relating to improved income and livelihoods. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 18.

Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was a significant amount of heterogeneity in the true outcomes (Q (1) =33.33, p=0.00, $\hat{\tau}^2$=0.31, I^2=93.92 per cent). Given that the meta-analysis only includes three studies, moderator analyses were not possible, but we conducted a sensitivity analysis to test if the variance in effect

sizes might be caused by factors related to the applied evaluation design (i.e. study type, risk of bias and period of follow-up). The GRADE framework shows the evidence on the effects of reminders on development results of improved income and livelihoods to be very low in quality (see Appendix 4).

Figure 18. Meta-analysis of reminders on development results: improved income and livelihoods



Source: Authors

Sensitivity analysis of reminders on development results: improved income and livelihoods

There are insufficient studies in the design categories to test the systematic influence of study design – all studies are RCTs and fall within the same period of follow-up. Further, the meta-analysis results are not systematically influenced by the risk of bias of the studies. Table 8 presents the results of the sensitivity analysis of reminders on development results in relation to improved income and livelihoods.

Table 8. Sensitivity analysis of reminders on development results: improved income and livelihoods

Variable	SMD	95% CI	Q	TAU²	I²	P-VALUE (Q)	SAMPLE	SENSITIVITY
Reminders: all studies	0.44	-0.21 to1.09	33.33	0.31	93.92	0.00	3	
Design	Lack of sufficient evidence in each category for sensitivity analysis							N/A
RCT	N/A						3	
Quasi-experimental design	N/A	No observations					0	
Risk of bias			1.10			0.29		Not sensitive
Low risk of bias	0.61	-0.33 to 1.56	29.26	0.45	96.58	0.00	2	
Moderate risk of bias	0.07	-0.30 to 0.11	0.00	0.00	.	.	1	

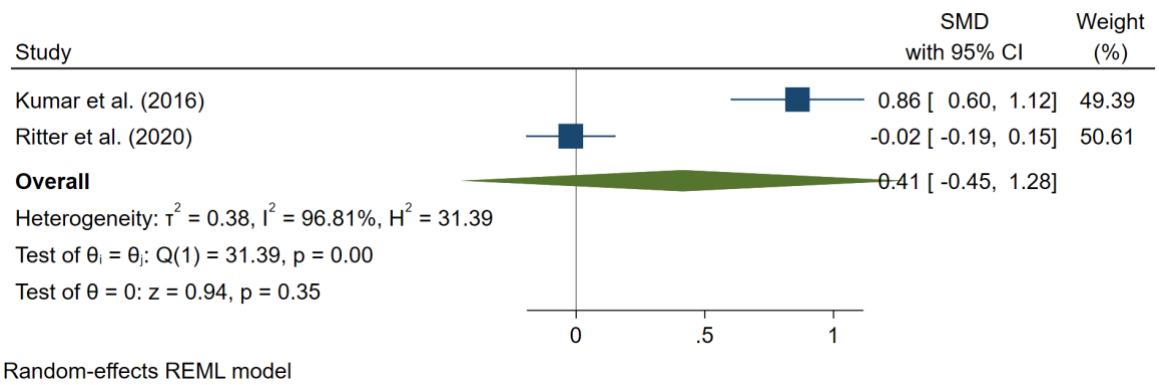
Variable	SMD	95% CI	Q	Tau^2	I^2	p-value (Q)	Sample	Sensitivity
High risk of bias	N/A	No observations						
Period of follow-up	Lack of sufficient evidence in each category for sensitivity analysis							N/A
Less than 1 month	0.44	-0.21 to 1.09	33.33	0.31	93.92	0.00	3	
1–6 months	N/A	No observations					0	
7–12 months	N/A	No observations					0	
13–18 months	N/A	No observations					0	
19–24 months	N/A	No observations					0	
>25 months	N/A	No observations					0	

Source: Authors

4. Effects of reminders on development results: improved health

We included k = 2 studies in the meta-analysis of the impact of reminders on development results for improved health. In terms of design, the analysis includes an RCT and a quasi-experimental design. One study is assessed as having a moderate risk of bias and the other as having a high risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.41 (CI: -0.45 to 1.28). The average outcome did not differ significantly from zero (z=0.94, p=0.35), indicating that while reminders present a positive overall effect, it is not statistically significant. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 19. Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was a significant amount of heterogeneity in the true outcomes (Q (1) =31.39, p=0.00 $\hat{\tau}^2$=0.38, I^2=96.81 per cent). With only two studies, sensitivity and moderator analyses were not possible. Applying GRADE indicates that the evidence on the effects of reminders on development results for improved health is of very low quality (see Appendix 4).

Figure 19. Meta-analysis of reminders on development results: improved health



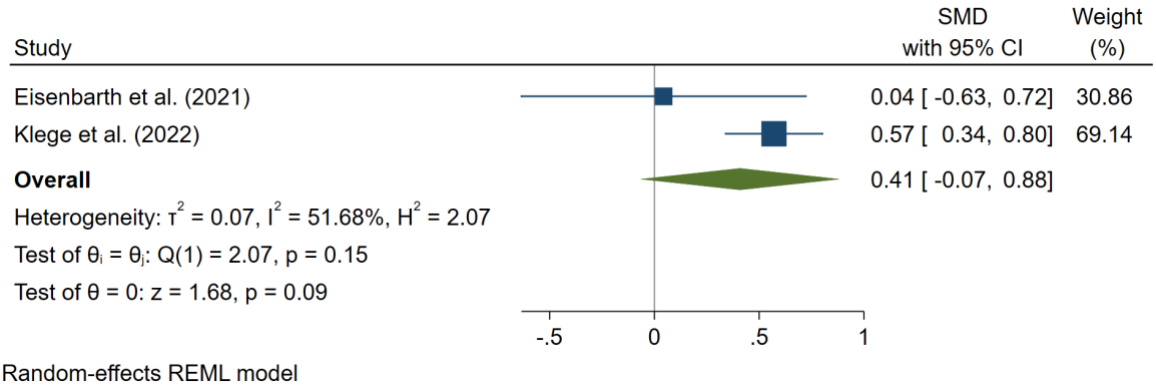
Source: Authors

5. EFFECTS OF REMINDERS ON BEHAVIOURAL OUTCOMES: FOREST USE AND ELECTRICITY CONSUMPTION

We included k = 2 studies in the analysis of the impact of reminders on behavioural outcomes in forest use and electricity consumption. Both studies are RCTs. One study is assessed as having a moderate risk of bias and the other as having a high risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.41 (CI: -0.07 to 0.88), and is therefore statistically insignificant (z=1.68, p=0.09) – that is, reminders have on average a significant positive effect on forest use and electricity consumption. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 20.

Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was no significant amount of heterogeneity in the true outcomes (Q (1) =2.07, p=0.15 $\hat{\tau}^2$=0.07, I^2=51.68 per cent). With only two studies, and given there was no heterogeneity among the effects, sensitivity and moderator analyses were not possible. Tests of publication bias are not valid. The GRADE framework shows that the evidence on the effects of reminders on behavioural outcomes for forest use and electricity consumption is of very low quality (see Appendix 4).

Figure 20. Meta-analysis of reminders on behavioural outcomes: forest use and electricity consumption



Source: Authors

6. NARRATIVE SYNTHESIS OF REMINDERS ON BEHAVIOURAL OUTCOMES

We also investigated whether a narrative synthesis of reminder interventions is feasible for interventions and outcome combinations not covered in the meta-analysis. However, due to heterogeneity in the intervention types, these studies could not be synthesized. This can be seen in Table 9, which provides a narrative overview of the individual studies and key findings on impact. More details on the interventions and the main findings on the effects are highlighted in Appendix 6.

Table 9. Narrative synthesis of reminders on behavioural outcomes interventions

STUDY	INTERVENTION TREATMENT COMPONENTS	CONTEXT	FINDINGS
Dzanku and others (2021) Low risk of bias	Reminders	Sub-Saharan Africa: Mali Income group: Low and middle income	There was improvement in uptake of storage technology (SMD: 0.2248, CI: -0.1877-0.6372). There was no significant improvement in crop sales as measured by gross revenue from the cereals traded ($) (SMD: 0.0119, CI: -0.3992 to0.4231).
Kumar and others (2016) High risk of bias	Reminders	South Asia: Pakistan Income group: Low and middle income	There was a significant statistical change of attitudes regarding infectious waste management before and after the intervention by comparing mean scores within the groups (SMD: 1.1671, CI: 0.8695 to 1.4647).
Manaseki-Holland and others (2021) Low risk of bias	Reminders	Sub-Saharan Africa: The Gambia Income group: Low-income country	There was a 4.4-fold increase in the practice of the five key behaviours (including but not limited to handwashing with soap and water before preparing food, washing of pots and utensils and drying them before use) in the intervention group at 6 months (IRR = 4.46 (CI: 3.63 to 5.47), p-value =0.001) when compared to controls. This significant effect was again observed at the 32-month assessment.
Shah and others (2020) Low risk of bias	Reminders	Latin America and Caribbean: Mexico Income group: Upper middle income	Study showed that treated participants were significantly (10% level) more likely to make a savings contribution and contributed more money overall in the 2-month period following the conclusion of the experiment (SMD: 0.0223, CI: -0.0012 to 0.0459).

Source: Authors

C. INTERVENTION III: GOAL SETTING

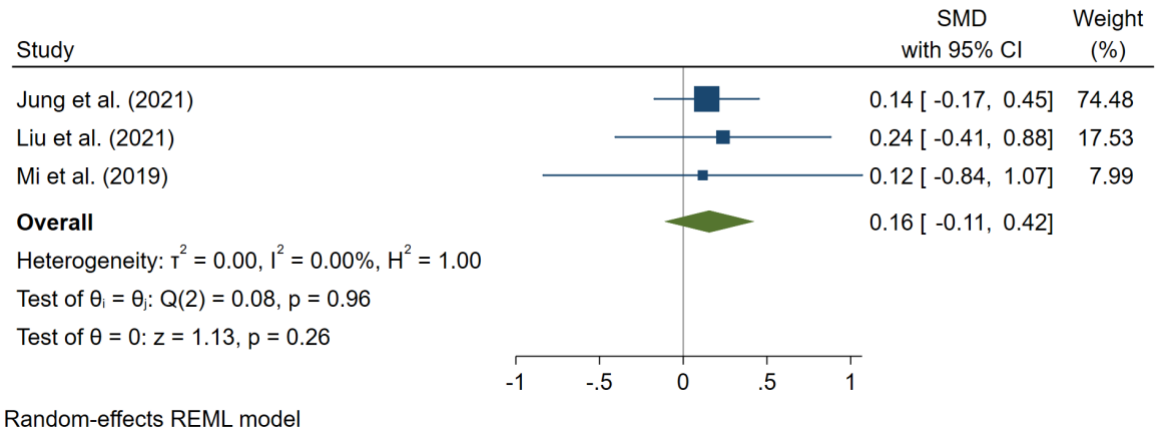
1. EFFECTS OF GOAL SETTING ON BEHAVIOURAL OUTCOMES: ELECTRICITY CONSUMPTION

We included $k = 3$ studies in the meta-analysis of goal setting on behavioural outcomes. This includes two RCTs and one quasi-experimental study. We assessed the three studies to be of low risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.16 (CI: -0.11 to 0.42), and therefore is not statistically significant (z=1.13, p<0.26). It can be said that goal setting as an intervention has a small positive but statistically insignificant effect on electricity consumption behaviour. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 21.

Given the small number of studies and the insignificance of the average outcome, this result should be interpreted with caution. Applying GRADE, evidence on the effects of goal setting on behavioural outcome in electricity consumption is very low in quality (see Appendix 4). According

to the Q-test, there is no significant amount of heterogeneity in the true outcomes (Q (1) =0.08, p=0.96, $\hat{\tau}^2$=0.00, I^2=0.00 per cent). Given the three studies did not display significant heterogeneity in the three studies, sensitivity and moderator analyses were not possible. Tests of publication bias are not valid.

Figure 21. Meta-analysis of goal setting on behavioural outcomes: electricity consumption



Source: Authors

Sensitivity analysis of goal setting on behavioural outcomes: electricity consumption

The results of the meta-analysis shown in Table 10 below indicate that quasi-experimental studies have larger effects, but this influence is not statistically significant. There is insufficient data in each category to performs sensitivity analysis on risk of bias and the period of follow-up, as all the studies are categorized as having low risk of bias and a less than 1 month follow-up period.

Table 10. Sensitivity analysis of setting on behavioural outcomes: electricity consumption

VARIABLE	SMD	95% CI CONSUMPTION	Q	TAU²	I²	P-VALUE (Q)	SAMPLE	SENSITIVITY
Reminders: all studies	0.16	-0.11 to 0.42	0.08	0.00	0.00	0.98	3	
Design			0.08			0.78		Not sensitive
RCT	0.14	-0.16 to 0.43	0.00	0.00	0.00	0.96	2	
Quasi-experimental design	0.24	-0.41 to 0.88	0.00	0.00	0.00	0.00%	1	
Risk of bias	Lack of sufficient evidence in each category for sensitivity analysis							N/A
Low risk of bias	0.16	-0.11 to 0.42	0.08	0.00	0.00%	0.98	3	
Moderate risk of bias	N/A	No observations	0.00	0.00	0.00%	0.00	0	
High risk of	N/A	No	0.00	0.00	0.00%	0.00	0	

VARIABLE	SMD	95% CI CONSUMPTION	Q	TAU^2	I^2	P-VALUE (Q)	SAMPLE	SENSITIVITY
bias		observations						
Period of follow-up	Lack of sufficient evidence in each category for sensitivity analysis							N/A
Less than 1 month	N/A	N/A	N/A	N/A	N/A	N/A	3	
1–6 months	N/A	No observations					0	
7–12 months	N/A	No observations					0	
13–18 months	N/A	No observations					0	
19–24 months	N/A	No observations					0	
>25 months	N/A	No observations					0	

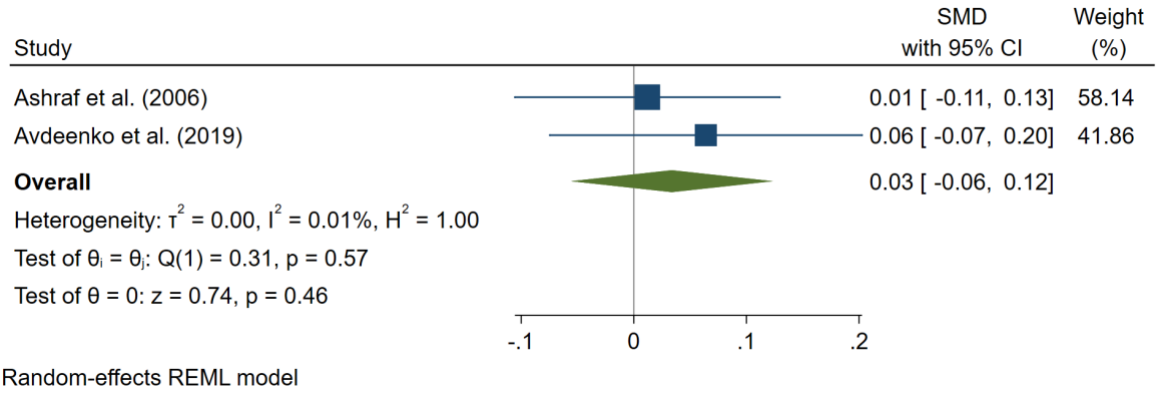
Source: Authors

2. EFFECTS OF GOAL SETTING AND COMMITMENT DEVICES ON BEHAVIOURAL OUTCOMES: SAVINGS BEHAVIOUR

We included k = 2 studies in the meta-analysis of the impact of goal setting and commitment devices on behavioural outcomes in savings behaviour. The two studies are both RCTs, assessed as having a low risk of bias. The pooled effect estimate based on the random-effects model was $\hat{\mu}$=0.03 (CI: −0.06 to 0.12), and is not statistically significant (z=0.74, p=0.46). This indicated that goal setting and commitment devices do lead to small positive but statistically insignificant effects on behavioural outcomes in savings behaviour. A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 22.

Given the small number of studies, this result should be interpreted with caution. According to the Q-test, there was no significant amount of heterogeneity in the true outcomes (Q (1) =0.31, p=0.57 $\hat{\tau}^2$=0.07, I^2=0.01 per cent). With only two studies, and given there was no heterogeneity among the effects, sensitivity and moderator analyses were not conducted. According to the GRADE framework, evidence on the effects of goal setting and commitment devices on behavioural outcomes in savings behaviour is rated very low in quality (see Appendix 4).

Figure 22. Meta-analysis of goal setting and commitment devices on behavioural outcomes: savings behaviour



Source: Authors

3. NARRATIVE SYNTHESIS OF GOAL-SETTING INTERVENTIONS

We also investigated whether a narrative synthesis of goal-setting interventions is feasible for interventions and outcome combinations not covered in the meta-analysis. However, due to heterogeneity in the intervention types, these studies could not be synthesized, as can be seen in Table 11. A narrative overview of the individual studies and key findings on impact are provided in Table 11. More details on the interventions and main findings on the effects are highlighted in Appendix 7.

Table 11. Narrative synthesis of goal-setting interventions

STUDY	INTERVENTION TREATMENT COMPONENTS	CONTEXT	FINDINGS
Ashraf and others (2005) Low risk of bias	Goal setting and commitment device	South Asia: Philippines Income group: Low and middle income	**Savings attitudes** In general, there was no significant impact on savings discipline (SMD: 0.1108, CI: -0.0075 to 0.2292), savings habit (SMD: 0.0439, CI: 0.0744 to 0.1622), and on savings practice in times of little cash (SMD: 0.1066, CI: - 0.0117 to 0.225). Further, treated participants reported feelings of regret about their spending practices and wished for a more disciplined savings habit (SMD: 0.3384, CI: 0.2194 to 0.4575).
Dalla and others (2021) Low risk of bias	Goal setting	Sub-Saharan Africa: South Africa Income group: Low and middle income	**Savings** The study demonstrated a significant increase in savings balances of participants in the goal card programme based on fortnightly collected data (SMD: 0.6394, CI: 0.4081 to 0.8706).
Grohmann and others (2020) Low risk of bias	Goal setting and planning prompts	Sub-Saharan Africa: Uganda	**Savings** The intervention did not influence the savings index among treated participants (SMD: 0.0621, CI: -0.1209 to 0.2451).

STUDY	INTERVENTION TREATMENT COMPONENTS	CONTEXT	FINDINGS
		Income group: Low income	
Kast and others (2018) Low risk of bias	Goal setting and micro-incentives	Latin America and Caribbean: Chile Income group: High income	**Monthly savings** There was an increase in the average monthly savings balances in the intervention group (SMD: 0.0565, CI: -0.049 to 0.1621). Evidence points to treated participants being encouraged to make frequent deposits and consequently raising their savings balances.
Liu and others (2021)	Goal setting and micro-incentives	East Asia and the Pacific: Singapore Income group: High income	**Electricity consumption** Households that set an electricity saving goal by themselves showed a significant reduction in electricity consumption (SMD: 0.2003, CI: -0.2565 to 0.6572) compared to the controls. **Water consumption** The study further observed a spill-over effect of self-set electricity saving intervention to a 15.1% reduction in water consumption among treated households (SMD: 0.3076, CI: -0.1477 to 0.763). However, there was no spill-over effect on water consumption among households that were assigned an electricity savings goal only (SMD: 0.1901, CI: -0.2636 to 0.6438) or even an additional micro-incentive (SMD: 0.1433, CI: -0.3129 to 0.5996).

Source: Authors

D. INTERVENTION IV: SALIENCE (EXPERIENCE DESIGN)

Overall, we find a limited amount of evidence on salience (communication) intervention that is heterogeneous at intervention level, but find two studies that have similar interventions and focus on the same outcome measures. However, we did not have sufficient evidence data to allow for a meta-analysis, as one study does not provide sufficient data for standardized effect size calculations.

Narrative synthesis on salience (experience design) on behavioural outcomes interventions

We then investigated whether a narrative synthesis of salience (experience design) interventions is feasible for interventions and outcome combinations in the absence of a meta-analysis. As a result, we conducted a narrative synthesis of two studies, namely Amon-Tanoh and others (2021) and Huang and others (2021).[10] Due to heterogeneity in the intervention types in the rest of the studies, these studies could not be synthesized, as can be seen in Table 12. A narrative overview of the individual studies and key findings on impact are also provided in the table. More details on the interventions and the main findings on the effects are highlighted in Appendix 8.

[10] The lack of sufficient data for converting the effect size to standardized mean difference in Amon-Tanoh and others (2021) ruled out a meta-analysis of the two studies.

Table 12. Narrative synthesis on salience (experience design) on behavioural outcomes interventions

STUDY	INTERVENTION TREATMENT COMPONENTS	CONTEXT	FINDINGS
Amon-Tanoh and others (2021) Low risk of bias	Salience (experience design) and reminders	Sub-Saharan Africa: Cote d'Ivoire Income group: Lower-middle-income country	At 1-month follow-up there was strong evidence of a large increase in the proportion of occasions in which handwashing with soap (HWWS) was done after using the toilet (risk ratio: 4·82, CI: 3·06 to 7·59, p<0·0001). The impact observed was largely sustained 5 months after intervention delivery (risk ratio: 2·68, CI: 1·65 to 4·34, p<0·0001).
Figueroa and others (2019) High risk of bias	Salience (experience design) and micro-incentives	Sub-Saharan Africa: Kenya Income group: Lower-middle-income country	At follow-up immediately after the intervention, the intervention resulted in higher uptake rates of energy-saving compact fluorescent lamp light bulbs (CFLs): SMD: 0.4389, CI: 0.1126 to 0.7653). The impact diminished at 2-year follow-up (SMD: 0.0958, CI: -0.2525 to 0.4441).
Huang and others (2021) Low risk of bias	Salience (experience design) and reminders	South Asia: Philippines Income group: Lower-middle-income country	The intervention increased the rate of student handwashing after toilet use: Handwashing rates after toilet use, with water and soap (SMD: 0.3532, 95%CI: -0.0439 to 0.7502). Handwashing rates after toilet use with at least water (SMD: 0.5326, CI: 0.1317 to 0.9335).
Tidwell and others (2020) Moderate risk of bias	Salience (experience design) and public commitments	Latin America and Caribbean: Colombia Income group: Upper-middle-income country	Children in the treatment reported handwashing with soap on key occasions (35.2%), which is higher than those in the control group (20.1%). (Risk ratio: 1.77, CI: 1.22 to 2.58, p =.003).
Yamin and others (2020) High risk of bias	Salience (experience design), salience (communication), social benchmarking and feedback	Latin America and Caribbean: Colombia Income group: Upper-middle-income country	Statistically significant decrease in average monthly fuel consumption (or increase in average distance travelled with one US gallon of fuel) at 1 month post-test, (SMD: 0.3882, CI: -0.0949 to 0.8714). At two months, the magnitude of impact declined (SMD: -0.2804, CI: -0.7616 to 0.2007) but increased in the third month (SMD: 0.4202, CI: -0.0636 to 0.9041).

Source: Authors

The Amon-Tanoh and others (2021) study that combined salience (experience design) and reminders, involved scripted videos depicting everyday scenes typical of compound life and incorporated comic elements and characteristics particular to popular local television series. The scripted videos depicted societal issues related to handwashing. The intervention also included posters designed by a local graphics artist based on the videos, that were placed on the inside of toilet doors and at the toilet's entrance and a handwashing station. The handwashing stations were also supplied with bottles of soapy water placed in easily visible positions. The study found

evidence of intervention effects at follow up both 1 month and 5 months later. At the 1-month follow-up, there was strong evidence of a large increase in the proportion of observed occasions in which handwashing with soap was done after using the toilet. This was observed in 143 (24 per cent) of 588 occasions (risk ratio: 4·82, CI: 3·06 to 7·59, p<0·0001). The change observed was largely sustained 5 months after intervention, and observed in 98 (22 per cent) of 450 occasions (risk ratio: 2·68, CI:1·65 to 4·34, p<0·0001).

The Huang and others (2021) study in the Philippines also aimed to promote student handwashing after toilet use through reminders and salience (experience design) in the form of contextual cues (painted footpaths, arrow stickers) pointing to the soap dish, and intended to trigger reactive processes to draw attention to the soap and therefore handwashing in the space. It also used other visual reminders such as posters and eyes stickers to serve as a reminder function, directly counteracting forgetfulness and present bias. The intervention was seen to increase the rate of student handwashing after toilet use. Specifically, the handwashing rate in treatment schools was 17.3 per cent (CI: 4.2 to 30.4) as compared to 11.7 per cent among students at control schools. The study used two measures for handwashing rates – after-toilet use with water and soap (SMD: 0.3532, CI: -0.0439 to 0.7502) and after-toilet use with at least water (SMD: 0.5326, CI: 0.1317 to 0.9335). Of note, is that the intervention had no differential handwashing impact by grade group or gender.

E. Intervention V: Salience (communication)

Overall, we find a limited amount of evidence on salience (communication) intervention that is heterogenous across interventions and outcomes to allow meta-analysis in this intervention type.

Narrative synthesis on salience (communication) on behavioural outcomes interventions

We investigated whether a narrative synthesis of salience (communication) is feasible for interventions and outcome combinations not covered in the meta-analysis. However, due to heterogeneity in the intervention types, these studies could not be synthesized, as can be seen in Table 13, which provides a narrative overview of the individual studies and key findings on impact. More details on the interventions and main findings on the effects are highlighted in Appendix 9.

Table 13. Narrative synthesis on salience (communication) on behavioural outcomes interventions

Study	Intervention treatment components	Context	Findings
Fiorillo and others (2014) Moderate risk of bias	Goal setting, salience (communication), reminders and commitment devices	South Asia: Philippines Income group: Middle income	Treatment group had savings balances that were 37 per cent higher than those in the control group after 8 weeks of account opening (SMD: 0.6598, CI: 0.4842 to 0.8355). This result was statistically significant across the full sample but not for sub-samples.
Miranda and others (2020) Low risk of bias	Goal setting, planning prompts and salience (communication)	Latin America and Caribbean: Costa Rica Income group: Upper middle	The study did not observe post-intervention effects on water consumption after 7 months (SMD: 0.0101, CI: 0.0637 to 0.0838).

Study	Intervention treatment components	Context	Findings
Wang and others (2018) Moderate risk of bias	Salience (communication), Commitment devices, micro-incentives	Sub-Saharan Africa: Uganda Income group: Low income	There was a significant increase in the log amount of savings at 24 months among the treated participants (combined SMD: 0.23835) at 0.1% significance level when compared to controls.
Wang and others (2021) Moderate risk of bias	Salience (communication), feedback, commitment devices and micro-incentives	Sub-Saharan Africa: Uganda Income group: Low income	The study demonstrated a significant decrease in poverty incidence rates from baseline to year 1, baseline to year 2 and baseline to year 3. However, although poverty incidence rates decreased from baseline to year 4 (SMD: 0.0585, CI: -0.0573 to 0.1744), this effect was not significant.
Young (2017) Moderate risk of bias	Salience (communication)	Sub-Saharan Africa: Malawi Income group: Low income	**Knowledge and awareness** The study did not find significant differences in knowledge of inter-relationships between sustainability issues with regard to climate change and human health (SMD: -0.1915, CI: -0.4103 to 0.0273). **Communication, information sharing and saving resources** There was a minimal increase in reports of communication among treated participants and controls (SMD: 0.1178, CI: -0.2695 to 0.5051). No treatment effects reported in information sharing – this was noticed only in control participants (SMD: -1.0185, CI: -1.3915 to -0.6454). Intervention influenced an increase in saving of resources in the treatment group (SMD: 0.3449, CI: 0.0298 to 0.6601).

Source: Authors

F. Subgroup analysis and investigation of heterogeneity

We conducted moderator analyses on a combination of extrinsic characteristics and substantive characteristics, namely: publication type and date, length of intervention, context (country, region, country income group), presence of incentives in the intervention, and implementing agency type. Our threshold for conducting moderator analysis referred to a minimum of four studies included in a given meta-analysis. Due to lack of sufficient evidence, moderator analysis was carried out in only two meta-analyses: the meta-analysis of feedback on water and electricity consumption (n = 5) and the meta-analysis of reminders on knowledge acquisition (n = 4). Across the two meta-analyses, the moderator variables were significant publication type and date, context (country, region, country income group), presence of incentives in the intervention, and implementing agency type. We only find a clear pattern in the influence of contextual characteristics – namely country, region and country income group – on the robustness of the meta-analysis, as these variables are significant in both meta-analysis and the robustness of the meta-analysis.

G. SENSITIVITY ANALYSIS

In conducting sensitivity analysis, we investigated whether the variance in effect sizes might be caused by factors related to the applied evaluation design (i.e. study type, risk of bias and period of follow-up). We conducted sensitivity analysis in six of the 12 meta-analyses where three or more studies are included in the meta-analysis.

Sensitivity analysis was possible in the following six meta-analyses: feedback on water and electricity consumption (n = 5), feedback and goalsetting on electricity consumption (n = 3), reminders on crop yield (n = 3), reminders on improved income and livelihoods (n = 3), reminders on knowledge acquisition (n = 4) and goal setting on electricity consumption (n = 3). In four of the six meta-analyses, it was not feasible to conduct sensitivity analysis on the study design variable as the meta-analysis includes either only RCTs or quasi-experimental design. Within the remaining moderators, there was no clear pattern identified. In the meta-analysis of feedback and impact on water and electricity consumption featuring three RCTs and two quasi-experimental designs, randomized trials have significantly the larger effect on the results of the pooled effect estimate. In the meta-analyzing of reminders on acquisition of knowledge, the pattern is different with three RCTs, and the quasi-experimental study has a significantly larger effect on the results of the meta-analysis.

In all but one of the (n = 5) meta-analyses we found that the results were robust to the risk of bias of the included studies. In the one meta-analysis where the risk of bias variable is sensitive, we see the effects from moderate risk of bias studies are significantly higher compared to low risk of bias studies (reminders on crop yield). Finally, regarding the follow-up variable within the sensitivity analysis, only two meta-analyses include sufficient evidence for sensitivity analysis (reminders on crop yield and reminders on knowledge acquisition).

The reminders on yield meta-analysis showed that studies with a follow-up of less than a month have larger effects than studies with 1–6 month follow-up. Lastly, in the study assessing reminders and knowledge, the results are sensitive to follow-up and the effects are higher for studies with a follow-up period of 1–6 months relative to a follow-up period of less than a month.

VI. DISCUSSION

A. OVERALL COMPLETENESS AND APPLICABILITY OF EVIDENCE

With a behavioural science lens, we can examine the results of this SR in terms of *types of behavioural interventions*, *types of behaviours*, and *relation to the theory of change*.

1. TYPES OF BEHAVIOURAL INTERVENTIONS

The most effective intervention according to this analysis is feedback, particularly in relation to electricity and water consumption. This is not surprising for two reasons. The first is that feedback is a very common behavioural intervention to influence a variety of behaviours, and widely studied in a variety of settings. Simply put, providing people with information about the results of their actions can reliably influence their behaviours. The second reason is that the mechanisms (both psychological and practical) behind feedback interventions are relatively straightforward. This means there may be less variation in how feedback interventions are applied, which in turn means that both intervention design and effects are less heterogeneous than they are for some other interventions. The upshot, for the purposes of this review, is that the evidence base for feedback is more consistent and rigorous compared to the other four types of interventions surveyed. It can also be argued that feedbacks are most effective because humans are social beings and this intervention invites interaction with other perspectives, whereas other interventions were used as non-social (despite being behavioural) mechanisms.

The story is similar for reminders, which we also find to be comparatively effective, even if the evidence base is of lower quality than for the feedback interventions. Because reminders aim to influence behaviour at very specific time points and are relatively simple to implement, intervention design and outcomes measurement can be more straightforward than for some other interventions. Whilst reminders are a very common tool in the behaviour modification toolkit, their relevance specifically to climate-change relevant behaviours appears to be weaker. The results from the studies in the meta-analysis – relating for example to crop yield and improved income and livelihoods – may be partially explained by the difficulty of connecting a reminder not only to a timely decision point but also to a specific behaviour relevant to climate change adaptation and mitigation. Both the challenge and the promise here are to do further work to identify relevant behaviours amenable to reminders, then to design appropriate interventions and evaluate them.

The goal setting interventions failed to demonstrate statistically significant effects. Given that goal setting is an oft-employed intervention in other contexts, with an extensive evidence base, we can only speculate as to the reasons behind the unconvincing results here. It may be that this kind of planning behaviour is most effective when the goals have significant personal meaning. The savings goals in the reviewed studies should theoretically meet that condition, but in broader perspective, it's possible that time-distant, climate-change related goals will be less personally resonant, and therefore have less power to motivate behaviours.

The remaining three behavioural interventions in the review – goal-setting, salience communication and salience experience design – do not provide robust evidence. With the latter two interventions, part of the complication may be that the general category of "salience" covers a very wide range of interventions as implemented. Though the general psychological principle of salience – that is, making something (whether a piece of information or a choice) stand out – is well established, the application of a salience "nudge" can take very different forms, with much more variation than, for example, reminders. That heterogeneity of intervention implementation means that both

categorizing and then attributing causal effects to intervention components is more complicated.[11] The very low number of studies and evidence available in the field and how scattered it is in terms of types of interventions, strengthens the importance of this work regarding the need to generate and use more of this type of evidence.

2. TYPES OF BEHAVIOURS

The types of behaviours examined in this review are limited. Those that appear in multiple studies include financial behaviour (here primarily saving money), reducing electricity/water consumption, and in the WASH [water, sanitation and hygiene] domain (here primarily handwashing). These specific behaviours were targeted with various types of interventions ranging from feedback to salience communication. As noted above, meta-analysis was not always possible due to extensive heterogeneity. But when it was, the analysis was conducted according to intervention type rather than behaviour type. Though the overall quality of evidence is weak, it is not possible to conclude from this review that certain types of interventions more effectively influence certain types of behaviours.

3. RELATION TO THE THEORY OF CHANGE

For a wide-angle view on these findings, it's worth returning to the theory of change outlined in the approach paper for this SR (section I.2 above). Below, we will consider how the behaviour categories from the theory of change relate to the meta-analysis. Here, though, we note how this review relates to the "development" section of the theory of change, and specifically its two subcategories: development sectors of behavioural intervention, and development results.

In terms of development sectors, the evidence relates primarily to energy and natural resource management, if we include water consumption in the latter. There is also some evidence relevant to the WASH sector, though handwashing is a narrow segment of that sector. This review does suggest that feedback and reminder interventions can impact energy and water consumption, but we are left with large gaps for the other sectors. Only a few studies in the review related to the agriculture sector, and those results were not robust. Education was even sparser, particularly if we conceive of this sector as *formal* education and not merely information provision.

Likewise, for the development results subcategory, the evidence from this review relates to only some sectors, namely supporting resource conservation, improving health, and improving income and livelihoods. Even in those cases, the evidence is often not robust, and causal links between specific behaviours and concrete development results can be difficult to establish. The evidence included in this review largely bypasses the other results sectors of enhancing equity, changing technologies, sustainable supply chain management and transport, and sustainable waste management.

B. QUALITY OF THE EVIDENCE

Based on an extensive search of both academic and grey literature sources, our SR identified a heterogeneous evidence base of behavioural science interventions in developing countries that includes 60 studies. Following the exclusion of 18 studies deemed to have a critical risk of bias, the largest number of studies with similar interventions and outcomes was five. The most frequently observed number of studies in a single intervention category (and associated components) was two

[11] Though we tried to narrow down salience interventions into two types, we found that the studies reviewed nonetheless depended on a diversity of intervention types. This became apparent only after the studies' inclusion in the review.

studies. Consequently, it is challenging to conduct an extensive synthesis of the effectiveness of interventions, as the evidence base is spread thinly across different interventions and outcomes. Our most rigorous synthesis comprised the effects of five feedback interventions.

In terms of the quality of the individual studies included in the review, the identified evidence base was of mixed quality. A small majority of 36 studies were judged to be of either low or moderate risk of bias while 24 studies were of either high or critical risk of bias. In order to formally assess the overall quality of the evidence base, we apply the GRADE framework, which combines the risk of bias rating of the included studies with an assessment of the consistency, precision and directness of the included evidence base. Applying the GRADE framework, we established that the overall quality of the evidence included in our 12 meta-analyses was low (see Appendix 4). Only the quality of evidence included in the meta-analysis on feedback-only interventions was of moderate quality. All other syntheses were based on very low-quality evidence (n = 11). In summary, this heterogeneous, low-quality nature of the evidence base limits the findings of our SR. We are only able to reach cautious conclusions regarding the evidence on feedback-only interventions; for all other intervention categories, the low quality of the evidence base limits the conclusions that can be drawn from the included studies.

C. LIMITATIONS AND POTENTIAL BIASES IN THE REVIEW PROCESS

Referring again to the theory of change, we see that the review examined a narrow set of behavioural interventions. The theory of change lays out five categories of behavioural intervention: *How* the choice is made, *Why* the choice is made, *Who* is making the choice, *When* the choice is made, and *Which* choices are available. The specific behavioural interventions analysed in the review fell either in the When (reminders and feedback) or Which (goal setting and salience) categories. Contextualising the findings against the results of our EGM, which included a broader range of behavioural interventions and outcomes, means that many other behavioural interventions were not considered in the review, even if they formed part of the EGM. These present an area for potential further synthesis and we must acknowledge that there are many other potentially promising behavioural interventions that we have not considered in this review.

The second set of limitations relates to the design of our SR and the process of conducting it. In terms of review design, our SR subscribed to an aggregative review approach aiming to use statistical meta-analysis to synthesize the results of a homogeneous body of interventions and outcomes. To attain this objective, the scope of the SR was narrow in terms of the included study designs and the synthesis question of interest, that is, which interventions work and to what effect. This narrow scope, by implication, excludes qualitative evidence on contextual factors that might mitigate intervention effects and causal pathways.

In conducting the SR, we applied a range of quality assurance mechanisms in order to limit potential biases to the research process introduced by the review team. Firstly, this review followed a two-stage approach that first involved the creation and publication of an EGM on the review topic. This aimed to mitigate against defining a review scope and question that did not meet stakeholders' priorities. Secondly, the SR design and process took place prior to conducting the review and was outlined in a protocol to ensure the transparency and replicability of the review (Booth and others, 2022c)). Thirdly, our SR was based on an exhaustive search effort of both academic and grey literature. Fourth, we applied a structured coding and risk of bias tool to assess the trustworthiness of the included studies and to extract relevant information for the synthesis in a transparent and consistent manner. Fifth, we assessed the quality of our meta-analyses using moderator and

sensitivity analyses. Lastly, throughout the review process we had guidance from a multi-disciplinary advisory group engaged through an in-depth co-production model.

In sum, we have confidence that we have applied reasonable measures to reduce the potential bias in the design and conduct of this SR.

D. AGREEMENTS AND DISAGREEMENTS WITH OTHER STUDIES OR REVIEWS

Above we alluded to the fact that the relevant prominence of feedback and reminder interventions was not surprising. These are two of the most-studied interventions in the behavioural toolkit. Some form of feedback, in particular, has often been employed in nudges to reduce energy consumption; there is a substantial body of literature in this area. That this SR finds feedback reasonably effective for the same goal is generally consistent with the wider literature. A point of disagreement, however, is this review's inability to establish positive results for feedback combined with social benchmarking, which is due to a lack of studies identified for this intervention combination suitable for synthesis. Social comparison of household energy usage is one of the most common forms this intervention takes, with multiple successful examples in different countries. This may be a context-dependent effect, however, as many of the similar studies were implemented in high-income countries.

Financial behaviour is an area that cut across several of the intervention types considered here, including salience communication, reminders, and goal setting. Financial behaviour also happens to be a major focus of behavioural science research and practice. Results from that research and practice are highly variable, with different intervention types as well as specific implementations achieving positive effects in one case and nothing in others. Our findings largely mirror that wider picture, for example with a reminders intervention achieving significant results in one study (Shah and others, 2020) whilst goal setting and a commitment device showed no significant effects in another (Ashraf and others, 2005). This mixed pattern points to the need for systematization of evidence and practice, particularly as relates to developing countries and environmental/climate change effects. Finally, though the evidence from goal-setting interventions considered for this review is not robust, there is a considerable literature documenting successful interventions in a variety of other domains.

VII. AUTHORS' CONCLUSIONS

A. IMPLICATIONS FOR PRACTITIONERS

Feedback interventions as a tool to influence resource consumption could and should receive continued consideration. One of the difficulties in determining the effectiveness of such interventions is measurement: can utility providers of electricity, water, or other relevant services provide the kind of granular data necessary for influential feedback? Presumably, basic demands of billing would ensure at least moderately adequate data, though that assumption might fail in developing, and/or in rural, off-grid areas.

Reminder interventions are promising for three reasons. First, they can be applied to almost anything; what is the desirable, socially positive behaviour people should be encouraged to undertake? Second, the implementation of reminders is relatively simple; at a minimum, all that's necessary is sending an SMS. Third, reminders can be combined with other behavioural insights to maximize message effectiveness; for instance, even a short SMS message could incorporate social norms with a framing device. The challenge with reminders, when it comes to applications relevant to climate change, may be identifying appropriate key decision points that can be "nudged". For example, what would be the desirable behaviours that should be encouraged, and when exactly should people be reminded to take those actions?

An implication relevant to both practitioners and researchers from this review concerns the duration of effects of behavioural interventions. In most of the studies we examined, effects in terms of behavioural change are generally absent past the six-month mark. That can result from study design (if data collection stopped at six months) or simply from the attenuation of intervention effects. Regardless, we need more precision here, and in particular, an investment in research and programme design to study the longer-term persistence of relevant behavioural change.

B. IMPLICATIONS FOR RESEARCH

A fundamental challenge with behavioural interventions is that they often seek to produce systemic change via micro-adjustments in individuals' choices. This challenge has underlined this SR too in the sense that the analysis required identifying individual behaviours and interventions that would contribute to climate change adaptation and mitigation. Indeed, more work remains to be done to relate behavioural mechanisms of change with broad-scale development results. As an example of a potential, ongoing research question in this vein, how can (a) specific components of an individual's identity influence their decisions that lead to behavioural changes, that in turn (b) contribute to development results that ultimately (c) lead to positive impacts in human wellbeing and socio-ecological systems development?

Besides more theoretical and empirical work linking micro-scale behavioural interventions with macro-scale development results, the evidence base for such behavioural interventions needs to continue to be built. In particular, this review suggests that the evidence base is rather thin when it comes to the pool of countries included in the review, that is, non-Annex I countries. The lion's share of the most rigorous behavioural insights work still takes place in North America, Western Europe, or Australia. A worthy investment for future research is to identify the interventions found to be most consistently effective in developing countries, including within clusters of countries which display similar characteristics (whether economic, social or political). Relatedly, an open question remains on what types of interventions most effectively influence which types of

behaviour. This review has identified evidence gaps and areas where evidence is relatively more robust, but all tied to intervention types. Learning more about how specific interventions (such as commitment devices) can influence specific behaviours – for instance, reducing waste – and how such interventions may work across contexts, is a daunting but vital research task going forward.

Finally, the conducted EGM including a broader body of evidence of 84 studies presents a resource to researchers to conduct further synthesis and analysis across different behavioural science interventions and outcomes in developing countries.

REFERENCES

Amon-Tanoh, Maud A., and others (2021). Effects of a social norm-based handwashing intervention including handwashing stations, and a handwashing station-only intervention on handwashing with soap in urban Côte d'Ivoire: a cluster randomised controlled trial. *The Lancet Global Health*, vol. 9, No. 12, e1707–e1718.

Ashraf, Nava, Dean Karlan, and Wesley Yin (2005). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, vol. 121, No. 2, pp. 635–672.

_____ (2010). Commit to saving: Don't stop until you save enough. Project summary. Innovations for Poverty Action. Available at https://www.bhub.org/project/commit-to-saving-dont-stop-until-you-save-enough/.

Balmford, Andrew, 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 (15 September).

Binet, Silvia, and others (2021). Independent evaluation of the adaptation portfolio and approach of the Green Climate Fund. Evaluation Report No. 9. 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, No. 34, pp. 20503–20510.

Booth, Samantha, and others (2022a). Evidence review on behavioural change in developing countries: Approach Paper (February). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund. Available at https://ieu.greenclimate.fund/sites/default/files/document/220510-egm-behavioural-change-approach-paper-top_0.pdf

_____(2022b). Behavioural science interventions within the development and environmental fields in developing countries: An evidence gap map. Learning paper (forthcoming). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund.

_____(2022c). Evidence review on behavioural science interventions in development and environmental fields in developing countries: Protocol (forthcoming). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund.

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

Brick, Kerri, Samantha De Martino, and Martine Visser (2017). Behavioural nudges for water conservation: Experimental evidence from Cape Town. Preprint.

Byerly, Hilary, and others (2018). Nudging pro-environmental behavior: evidence and opportunities. *Frontiers in Ecology and the Environment*, vol. 16, No. 3, pp. 159–168.

Campbell Collaboration (2019). Campbell collaboration systematic reviews: Policies and guidelines. Available at https://doi.org/10.4073/cpg.2016.1.

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.

Center for Behaviour & the Environment (2020). Leveraging Social Norms to Reduce Household Energy Use. Available at https://behavior.rare.org/wp-content/uploads/2020/07/Social-Influences.-Opower.7.8.pdf.

Chance, Zoe, and others (2014) Why choosing healthy foods is hard, and how to help: presenting the 4Ps framework for behavior change. *Customer Needs and Solutions* 1, No. 4, pp. 253-262.

Chapman, Gretchen, and others (2016). Goals and social comparisons promote walking behaviour. *Medical Decision Making*, vol. 35, No. 4, pp. 472–478.

Cinner, Joshua (2018). How behavioral science can help conservation. *Science*, vol. 362, No. 6417, pp. 889–890.

Cochran, William G. (1954). The combination of estimates from different experiments. *Biometrics* vol. 10, No. 1, pp. 101-129.

Dalla Pellegrina, Lucia, and others (2021). Fostering savings by commitment: Evidence from a quasi-natural experiment at The Small Enterprise Foundation in South Africa. *World Development*, vol. 148.

Datta, Saugato, and Sendhil Mullainathan (2014). Behavioral design: A new approach to development policy. *Review of Income and Wealth*, vol. 60, No. 1, pp. 7–35.

De La Rue, Lisa, and others (2014). School-based interventions to reduce dating and sexual violence: a systematic review. *Campbell Systematic Reviews*, vol. 10, No. 1, pp. 1–110.

de Leon Barido, Diego Ponce, and others (2018). Opportunities for behavioural energy efficiency and flexible demand in data-limited low-carbon resource constrained environments. *Applied Energy*, vol. 228, pp. 512–523.

DerSimonian, Rebecca, and Nan Laird (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, vol. 7, No. 3, pp. 177-188.

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.

Dzanku, Fred Mawunyo, Robert Osei, and Isaac Osei-Akoto (2021). The impact of mobile phone voice message reminders on agricultural outcomes in Mali. *Agricultural Economics*, vol. 52, No. 5, pp. 789–806.

Edenhofer, Ottmar, and others, eds. (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.* Cambridge, New York: Cambridge University Press.

Ellis, Paul D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, New York: Cambridge University Press.

Figueroa, Aurelia, and others (2019). Show me (more than) the money! Assessing the social and psychological dimensions to energy efficient lighting in Kenya. *Energy Research and Social Science*, vol. 47, pp. 224–232.

Fiorillo, Alexandra, and others (2014). Applying behavioural economics to improve micro-savings outcomes. White paper. New York, Washington DC: ideas42, Grameen Foundation. Available at http://www.ideas42.org/wp-content/uploads/2015/05/Applying-BE-to-Improve-Microsavings-Outcomes-1.pdf

Flanagan, Ann Elizabeth, and Jeffery Clark Tanner (2016). Evaluating behaviour change in international development operations: A new framework. World Bank IEG Working Papers. Washington, DC: World Bank.

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, No. 6, pp. 801–827.

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

Goette, Lorenz, Ching Leong, and Neng Qian (2019). Motivating household water conservation: A field experiment in Singapore. *PLoS One*, vol. 14, No. 3, e0211891.

Gough, David, Sandy Oliver, and James Thomas (2017). *An introduction to systematic reviews*. Sage.

Grohmann, Antonia, Tabea Lakemann and Helke Seitz (2020). Savings goal calendars as soft commitment devices: Evidence from small business owners in Uganda. Discussion Paper No. 1919. Berlin: German Institute for Economic Research.

Hallsworth, Michael, and others (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, vol. 148, No. 4, 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, No. 7, pp.358–364.

Haynes, Laura, and others (2012). Test, learn adapt: Developing public policy with randomised controlled trials. Report. London: Behavioural Insights Team, Cabinet Office.

Hedges, Larry V. (2009). Statistical considerations. In *The handbook of research synthesis and meta-analysis*, 2nd ed., Harris Cooper, Larry V. Hedges, and Jeffrey C. Valentine, eds., pp 37-48. New York: Russel Sage Foundation. Available at https://www.jstor.org/stable/10.7758/9781610441384.

Hedges, Larry V., and Ingram Olkin (1985). *Statistical methods for meta-analysis.* Orlando, Florida: Academic Press.

Higgins, Julian P.T., and others (2003). Measuring inconsistency in meta-analyses. *BMJ*, vol. 327, No. 7414, pp. 557–560.

Higgins, Julian P.T., and others, eds. (2019). *Cochrane handbook for systematic reviews of interventions*. 2nd edition. Chichester: John Wiley & Sons.

Higgins, Julian P.T., and Simon Thompson (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, vol. 21, No. 11, pp. 1539–1558.

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.

Huang, Haijing Crystal, and others (2021). Nudging handwashing among primary school students in the Philippines: Evidence from a cluster randomized trial. *The American Journal of Tropical Medicine and Hygiene*, vol. 105, No. 6, pp. 1806–1815.

Jones, Nathalie D., and Benjamin D. Rosenberg (2018). Program theory of change. In *The SAGE encyclopedia of educational research, measurement, and evaluation*, Bruce B. Frey, eds, vol. 1, pp. 1315−1318. SAGE Publications, Inc. Available at https://sk.sagepub.com/reference/sage-encyclopedia-of-educational-research-measurement-evaluation/i16701.xml.

Karlan, Dean, and others (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, vol. 62, No. 12, pp. 3393–3411.

Kast, Felipe, Stephan Meier, and Dina Pomeranz (2018). Saving more in groups: Field experimental evidence from Chile. *Journal of Development Economics*, vol. 133, pp. 275–294.

Keef, Stephen P., and Leigh A. Roberts (2004). The meta-analysis of partial effect sizes. *British Journal Mathematical and Statistical Psychology*, vol. 57, No. 1, pp. 97–129.

Klege, Rebecca, and others (2022). The power of nudging: Using feedback, competition, and responsibility assignment to save electricity in a non-residential setting. *Environmental and Resource Economics*, vol. 81, pp. 573–589.

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.

Kumar, Ramesh, Ratana Somrongthong, and Jamil Ahmed (2016). Impact of waste management training intervention on knowledge, attitude and practices of teaching hospital workers in Pakistan. *Pakistan Journal of Medical Sciences*, vol. 32, No. 3, pp. 705-710.

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.

Lawlor, Blaine, and Martin 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.

Liu, Y., H.W. Kua, and Y. Lu (2021). Spillover effects from energy conservation goal setting: A field intervention study. *Resources, Conservation and Recycling*, vol. 170, No. 105570.

Locke, Edwin (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology*, vol. 5, No. 2, pp. 117–124. Available at https://www.sciencedirect.com/journal/applied-and-preventive-psychology.

Locke, Edwin, 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, No. 1, pp. 1–111.

Manaseki-Holland, Semira, and others (2021). Effects on childhood infections of promoting safe and hygienic complementary-food handling practices through a community-based programme: A cluster randomised controlled trial in a rural area of The Gambia. *PloS Medicine*, vol. 18, No. 1.

Miranda, Juan J., Saugato Datta, and Laura Zoratto (2020). Saving water with a nudge (or two): evidence from Costa Rica on the effectiveness and limits of low-cost behavioural interventions on water use. *The World Bank Economic Review*, vol. 34, No. 2, pp. 444–463.

Moher, David, and others (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (Chinese edition). *Journal of Chinese Integrative Medicine*, vol. 7, No.9, pp. 889-896.

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. 1, pp. 1–13.

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, No. 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, No. 1, pp. 1–14.

Polanin, Joshua R, and others (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, vol. 86, No. 1, pp. 207-236.

Rebelo Da Silva, Natalie, and others (2017). Making the most of what we already know: A three-stage approach to systematic reviewing. *Evaluation review*, vol. 41, No.2, pp. 155-172.

Ruiz-Tagle, J.C., and Alejandra Schueftan (2021). Nudging for cleaner air: Experimental evidence from an RCT on wood stove usage. *Environmental and Resource Economics*, vol. 79, No. 4, pp. 713–743.

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 Geert Verbong (2016). The roles of users in shaping transitions to new energy systems. *Nature Energy*, vol. 1, No. 5, pp. 1–7.

Service, Owain, and others (2014). EAST: Four simple ways to apply behavioural insights. The Behavioural Insights Team. Available at https://www.bi.team/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf.

Shah, Avni, and others (2020). Can making family salient increase retirement savings? Evidence from a large-scale field experiment. Working paper. Available at https://ssrn.com/abstract=3460722.

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, No. 3.

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, New York: Cambridge University Press.

Stern, Paul C. (2020). A reexamination on how behavioral interventions can promote household action to limit climate change. *Nature communications*, vol. 11, No. 1, 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, pp. 1–6.

Sterne, Jonathan A. C., and others (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, vol. 355(i4919), pp. 1–7.

Stewart, Lesley A., and others (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *Jama*, vol. 313, No. 16, pp. 1657-1665.

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, No. 3, pp. 355–366.

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

Thondhlana, G., and H.W. Kua (2016). Promoting household energy conservation in low-income households through tailored interventions in Grahamstown, South Africa. *Journal of Cleaner Production*, vol. 131, pp. 327–340.

Tidwell, James B., and others. (2020). Impact of a teacher-led school handwashing program on children's handwashing with soap at school and home in Bihar, India. *PLoS One*, vol. 15, No. 2.

Torres, Mónica M.J., and Fredrik Carlsson (2016). Social norms and information diffusion in water-saving programs: Evidence from a randomized field experiment in Colombia. Working paper in economics, No 652. Göteborg: University of Gothenburg.

United Nations Environment Programme (2017). *The emissions gap report 2017*. Synthesis report. Nairobi: United Nations Environment Programme (UNEP). 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.

Waddington, Hugh, and others (2012). How to do a good systematic review of effects in international development: A tool kit. *Journal of development effectiveness*, vol. 4, No. 3, pp. 359-387.

Wade, Keith, and Marcus Jennings (2015). Climate change and the global economy: Regional effects, 26 July. Available at https://www.schroders.com/en/ch/asset-management/insights/economics/climate-change-and-the-global-economy-regional-effects/.

Wang, Julia S.H., and others (2018). Effects of financial incentives on saving outcomes and material well-being: evidence from a randomized controlled trial in Uganda. *Journal of Policy Analysis and Management*, vol. 37, No. 3, pp. 602–629.

Wang, Julia S.H., and others (2021). A multifaceted intervention with savings incentives to reduce multidimensional child poverty: Evidence from the Bridges Study (2012–2018) in rural Uganda. *Social Indicators Research*, vol. 158, No. 3, pp. 947–990.

Williamson, Katie, and others (2018). *Climate change needs behavior change: Making the case for behavioral solutions to reduce global warming*. Arlington, VA: Rare. Available at https://rare.org/report/climate-change-needs-behavior-change/.

Wilson, David B., David Weisburd, and David McClure (2011). Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. *Campbell Systematic Reviews*, vol. 7, No. 1, pp. 1–53.

Yamin, Paulius, and others (2020). Local determinants of driving behaviours: Installation Theory interventions to reduce fuel consumption among truck drivers in Colombia. *Transport*, vol. 35, No. 6, pp. 616–634.

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, No. 52, e18353.

Studies included in the systematic review

Agarwal, Shilpa, Valentina Brailovskaya, and Jonathan Robinson (2020). Saving for multiple financial needs: Evidence from lockboxes and mobile money in Malawi. *The Review of Economics and Statistics*, pp. 1–45.

Agarwal, Sumit, and others (2017a). Nudges from school children and electricity conservation: Evidence from the "Project Carbon Zero" campaign in Singapore. *Energy Economics*, vol. 61, pp. 29–41.

Agarwal, Sumit, and others (2017b). *The role of goals and real-time feedback in resource conservation: Evidence from a large-scale field experiment.* Singapore: National University of Singapore. Available at https://www.iame.uni-bonn.de/people/lorenz-goette/goals-and-real-time-feedback.

Aguinaga, Paulina, and others (2019). Raising achievement among microentrepreneurs: An experimental test of goals, incentives, and support groups in Medellin, Colombia. *Journal of Economic Behavior and Organization*, vol. 161, pp. 79–97.

Aker, Jenny C., and others (2020). Just a bit of cushion: The role of a simple savings device in meeting planned and unplanned expenses in rural Niger. *World Development*, vol. 128.

Amon-Tanoh, Maud A., and others (2021). Effects of a social norm-based handwashing intervention including handwashing stations, and a handwashing station-only intervention on handwashing with soap in urban Côte d'Ivoire: a cluster randomised controlled trial. *The Lancet Global Health*, vol. 9, No. 12, e1707–e1718.

Apisarnthanarak, Anucha, Thanee Eiamsitrakoon, and Linda M. Mundy (2015). Behavior-based interventions to improve hand hygiene adherence among intensive care unit healthcare workers in Thailand. *Infection Control and Hospital Epidemiology*, vol. 36, No. 5, pp. 517–521.

Ashraf, Nava, Dean Karlan, and Wesley Yin (2005). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, vol. 121, No. 2, pp. 635–672.

_____(2010). Female empowerment: Impact of a commitment savings product in the Philippines. *World Development*, vol. 38, No. 3, pp. 333–344.

Aulady, Mohamad F.N., and Toshio Fujimi (2021). Reminders for learning about flood risk preparedness in Indonesian poor villages: Parallel group in situ randomized controlled trial. *Natural Hazards Review*, vol. 22, No. 3.

Avdeenko, Alexandra, Albrecht Bohne, and Markus Frölich (2019). Linking savings behavior, confidence and individual feedback: A field experiment in Ethiopia. *Journal of Economic Behavior and Organization*, vol. 167, pp. 122–151.

Blumenstock, Joshua Evan, Michael Callen, and Tarek Ghani (2016). Mobile-izing savings with automatic contributions: Experimental evidence on present bias and default effects in Afghanistan. Discussion paper. London: London Centre for Economic Policy Research.

Brick, Kerri, Samantha De Martino, and Martine Visser (2017). Behavioural nudges for water conservation: Experimental evidence from Cape Town. Preprint.

Burns, Justine, Brendan Maughan Brown, and Âurea Mouzinho (2018). Washing with hope: Evidence of improved handwashing among children in South Africa from a pilot study of a novel soap technology. *BMC Public Health*, vol. 18, No. 1, pp. 1–13.

Carrión-Yaguana, Vanessa D., Jeffrey Alwang, and Victor H. Barrera (2020). Promoting behavioural change using text messages: A case study of blackberry farmers in Ecuador. *Journal of Agricultural and Applied Economics*, vol. 52, No. 3, pp. 398–419.

Casaburi, Lorenzo, and others (2014). Harnessing ICT to increase agricultural production: Evidence from Kenya. Working paper. Cambridge, MA: Harvard University.

Chong, Aalberto, and others (2015). (Ineffective) messages to encourage recycling: Evidence from a randomized evaluation in Peru. *The World Bank Economic Review*, vol. 29, No. 1, pp. 180–206.

Cole, Shawn Allen, Mukta Joshi, and Antoinette Schoar (2021). Heuristics on call: The impact of mobile phone-based business management advice. Working paper. Available at https://ssrn.com/abstract=3971013.

Coville, Aidan, and others (2019). The Nollywood nudge: An entertaining approach to saving. Policy Research Working Paper No. 8920. Washington, D.C.: World Bank.

Dalla Pellegrina, Lucia, and others (2021). Fostering savings by commitment: Evidence from a quasi-natural experiment at The Small Enterprise Foundation in South Africa. *World Development*, vol. 148.

de Leon Barido, Diego Ponce, and others (2018). Opportunities for behavioural energy efficiency and flexible demand in data-limited low-carbon resource constrained environments. *Applied Energy*, vol. 228, pp. 512–523.

Dzanku, Fred Mawunyo, Robert Osei, and Isaac Osei-Akoto (2021). The impact of mobile phone voice message reminders on agricultural outcomes in Mali. *Agricultural Economics*, vol. 52, No. 5, pp. 789–806.

Eisenbarth, Sabrina, Louis Graham, and Anouk S. Rigterink (2021). Can reminders of rules induce compliance? Experimental evidence from a common pool resource setting. *Environmental and Resource Economics*, vol. 79, No. 4, pp. 653–681.

Figueroa, Aurelia, and others (2019). Show me (more than) the money! Assessing the social and psychological dimensions to energy efficient lighting in Kenya. *Energy Research and Social Science*, vol. 47, pp. 224–232.

Fiorillo, Alexandra, and others (2014). Applying behavioural economics to improve micro-savings outcomes. White paper. New York, Washington DC: ideas42, Grameen Foundation. Available at http://www.ideas42.org/wp-content/uploads/2015/05/Applying-BE-to-Improve-Microsavings-Outcomes-1.pdf.

Freeman, Matthew C., and others (2020). Impact of a demand-side integrated WASH and nutrition community-based care group intervention on behavioural change: a randomised controlled trial in western Kenya. *BMJ Global Health*, vol. 5, No. 11.

Goette, Lorenz, Ching Leong, and Neng Qian (2019). Motivating household water conservation: A field experiment in Singapore. *PloS One*, vol. 14, No. 3.

Grohmann, Antonia, Tabea Lakemann and Helke Seitz (2020). Savings goal calendars as soft commitment devices: Evidence from small business owners in Uganda. Discussion Paper No. 1919. Berlin: German Institute for Economic Research.

Huang, Haijing Crystal, and others (2021). Nudging handwashing among primary school students in the Philippines: Evidence from a cluster randomized trial. *The American Journal of Tropical Medicine and Hygiene*, vol. 105, No. 6, pp. 1806–1815.

Jung, Miyeon, Daegon Cho, and Euncheol Shin (2021). Repairing a cracked mirror: The heterogeneous effect of personalized digital nudges driven by misperception. *Production and Operations Management*, vol. 30, No. 8, pp. 2586–2607.

Karlan, Dean, and others (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, vol. 62, No. 12, pp. 3393–3411.

Kast, Felipe, Stephan Meier, and Dina Pomeranz (2018). Saving more in groups: Field experimental evidence from Chile. *Journal of Development Economics*, vol. 133, pp. 275–294.

Klege, Rebecca, and others (2022). The power of nudging: Using feedback, competition, and responsibility assignment to save electricity in a non-residential setting. *Environmental and Resource Economics*, vol. 81, pp. 573–589.

Kua, H.W., and S.E. Wong (2012). Lessons for integrated household energy conservation policies from an intervention study in Singapore. *Energy Policy*, vol. 47, pp. 49–56.

Kumar, Ramesh, Ratana Somrongthong, and Jamil Ahmed (2016). Impact of waste management training intervention on knowledge, attitude and practices of teaching hospital workers in Pakistan. *Pakistan Journal of Medical Sciences*, vol. 32, No. 3, pp. 705-710.

Larochelle, Catherine, and others (2019). Did you really get the message? Using text reminders to stimulate adoption of agricultural technologies. *The Journal of Development Studies*, vol. 55, No. 4, pp. 548–564.

Li, Zhaungai, and Xia Cao (2021). Analysis of information feedback on residential energy conservation and the implications: The case of China. *Frontiers in Environmental Science*, vol. 9.

Liu, X., and others (2019). A multimodal intervention to improve hand hygiene compliance via social cognitive influences among kindergarten teachers in China. *PLoS One*, vol. 14, No. 5.

Liu, Y., H.W. Kua, and Y. Lu (2021). Spillover effects from energy conservation goal setting: A field intervention study. *Resources, Conservation and Recycling*, vol. 170, No. 105570.

Luoto, Jill, and others (2014). Nudging to use: Achieving safe water behaviors in Kenya and Bangladesh. *Journal of Development Economics*, vol. 110, pp. 13–21.

Manaseki-Holland, Semira, and others (2021). Effects on childhood infections of promoting safe and hygienic complementary-food handling practices through a community-based programme: A cluster randomised controlled trial in a rural area of The Gambia. *PloS Medicine*, vol. 18, No. 1.

Mi, Lingyun, and others (2019). Using goal and contrast feedback to motivate Chinese urban families to save electricity actively – A randomized controlled field trial. *Journal of Cleaner Production*, vol. 226, pp. 443–453.

Miranda, Juan J., Saugato Datta and Laura Zoratto (2020). Saving water with a nudge (or two): evidence from Costa Rica on the effectiveness and limits of low-cost behavioural interventions on water use. *The World Bank Economic Review*, vol. 34, No. 2, pp. 444–463.

Naluonde, Tabonga, and others (2018). A disruptive cue improves handwashing in school children in Zambia. *Health Promotion International*, vol. 34, No. 6, pp. e119–e128.

Nourmoradi, Heshmatollah, and others (2021). The influence of an education program on students' environmental responsibility in developing countries: Evidence from Iran. *Journal of Human Behavior in the Social Environment*, pp. 1–11.

Pandur, Vlad, and Patricia Albulescu (2016). How to promote pro-environmental behaviour in organizations: The message matters. *Psihologia Resurselor Umane*, vol. 14, No. 1, pp. 69–82.

Rebelo Da Silva, Natalie, and others (2017). Making the most of what we already know: A three-stage approach to systematic reviewing. *Evaluation review*, vol. 41, No.2, pp. 155-172.

Ritter, Michael, and others (2020). Impact evaluation of an SMS campaign to promote household chlorination in rural Haiti. *Water*, vol. 12, No. 11.

Ruiz-Tagle, J.C., and Alejandra Schueftan (2021). Nudging for cleaner air: Experimental evidence from an RCT on wood stove usage. *Environmental and Resource Economics*, vol. 79, No. 4, pp. 713–743.

Shah, Avni, and others (2020). Can making family salient increase retirement savings? Evidence from a large-scale field experiment. Working paper. Available at https://ssrn.com/abstract=3460722.

Sharma, Upasna, and others (2021). Do phone-based short message services improve the uptake of agri-met advice by farmers? A case study in Haryana, India. *Climate Risk Management*, vol. 33.

Soman, Dilip, and Amar Cheema (2011). Earmarking and partitioning: Increasing saving by low–income households. *Journal of Marketing Research*, vol. 48(SPL), pp. S14–S22.

Stewart, Lesley A., and others (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *Jama*, vol. 313, No. 16, pp. 1657-1665.

Sudarshan, Anant (2017). Nudges in the marketplace: The response of household electricity consumption to information and monetary incentives. *Journal of Economic Behavior and Organization*, vol. 134, pp. 320–335.

Sukor, Nur S.A., Nur K. Basri, and Sitti A. Hassan (2015). Travel behaviour modification technique – A framework for awareness on carbon footprint towards adolescents in developing country. *Jurnal Teknologi*, vol. 76, No. 14.

Thondhlana, G., and H.W. Kua (2016). Promoting household energy conservation in low-income households through tailored interventions in Grahamstown, South Africa. *Journal of Cleaner Production*, vol. 131, pp. 327–340.

Tidwell, James B., and others. (2020). Impact of a teacher-led school handwashing program on children's handwashing with soap at school and home in Bihar, India. *PLoS One*, vol. 15, No. 2.

Torres, Mónica M.J., and Fredrik Carlsson (2016). Social norms and information diffusion in water-saving programs: Evidence from a randomized field experiment in Colombia. Working paper in economics, No. 652. Göteborg: University of Gothenburg.

Usmani, Faraz, Jason Steele, and Marc Jeuland (2017). Can economic incentives enhance adoption and use of a household energy technology? Evidence from a pilot study in Cambodia. *Environmental Research Letters*, vol. 12, No. 3.

Visser, Martine and others (2021). Saving water at Cape Town schools by using smart metering and behavioural change. *Water Resources and Economics*, vol. 34.

Wang, Julia S.H., and others (2018). Effects of financial incentives on saving outcomes and material well-being: evidence from a randomized controlled trial in Uganda. *Journal of Policy Analysis and Management*, vol. 37, No. 3, pp. 602–629.

Wang, Julia S.H., and others (2021). A multifaceted intervention with savings incentives to reduce multidimensional child poverty: Evidence from the Bridges Study (2012–2018) in rural Uganda. *Social Indicators Research*, vol. 158, No. 3, pp. 947–990.

Wichaidit, Wit, and others (2020). Installation of pedal-operated alcohol gel dispensers with behavioural nudges and changes in hand hygiene behaviors during the COVID-19 pandemic: A hospital-based quasi-experimental study. *Journal of Public Health Research*, vol. 9, No. 4.

Yamin, Paulius, and others (2020). Local determinants of driving behaviours: Installation Theory interventions to reduce fuel consumption among truck drivers in Colombia. *Transport*, vol. 35, No. 6, pp. 616–634.

Young, Carrie E. (2017). The impact of narrative and participatory drama on social interactions, attitudes, and efficacy around health and environmental issues in Malawi. PhD diss. Ithaca, NY: Cornell University.

APPENDICES

Appendix 1. DESCRIPTION OF INTERVENTION TYPES

This SR analyses the impact of feedback, reminders, salience (communication), salience (experience design), and goal setting interventions in attaining desired environmental and development outcomes in developing countries. Detailed descriptions of these intervention types are as follows:

Reminders. Reminders involve prompting people in a timely way to call their attention to something and encourage them to take certain actions. Reminders typically involve sending a message (whether digitally such as an email or SMS or via a letter) with a specific call to action (such as attending an appointment or filing taxes) at a timely moment (e.g. the day before an appointment or on the day that a submission needs to be made). Reminders help counteract one of the cognitive limitations that we face as human beings – the fact that we have limited attention and memory. Even if a behaviour or task to be completed is important (like filing taxes or going to the doctor), it may be forgotten because it competes for our attention with everything else that we need to do. Reminders anticipate that people are likely to forget and help direct attention to an action that needs to be taken, making it more salient and bringing it to the "top of mind" at the right moment.

Reminders can be delivered as "pure" reminders that remind someone to do something at the right time point. They can also be coupled with other behavioural interventions such as using gain or loss framing or including social norms in the messaging. They can be delivered as once-off reminders for once-off behaviours like attending an appointment, or 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, "pure" SMS reminders sent by the UK Courts service at the final point at which individuals could pay fines doubled fine payment amounts. 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 or outcome, and may include past performance or be in relation to others' behaviours. 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, feedback interventions can encourage continued progress. 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 (using too much electricity, for example). Feedback interventions can, however, backfire for those already performing well in relation to others (e.g. someone who learns that they are using less electricity than their neighbours may actually increase 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, decrease energy use, and increase recycling (Center for Behaviour & the Environment, 2020). For example, 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 & 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 content of messaging. They are distinct from reminders which focus on timely delivery. Simply sending out a communication (e.g. sending a letter or email) cannot be classified 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 simply 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 experience processes or interactions and not on messaging strategy.

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 increase likelihood of thorough completion, 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 that 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 provide clearly defined motivation to carry out 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, 2016). For example, a University in North America increased physical activity among staff members through a goal-setting intervention in which staff members were given daily step goals. Those receiving a high goal[12] walked an average of 1,912 more steps per day than those given a low goal (Chapman and others, 2016).

[12] University staff members received a low, medium, or high walking goal (10 per cent, 50 per cent, or 100 per cent increase over baseline walking).

Appendix 2. DATA EXTRACTION TOOL

The data extraction tool can be accessed in the approach paper for the evidence review and the SR protocol:

Booth, Samantha, and others (2022a). Evidence review on behavioural change in developing countries: Approach Paper (February). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund. Available at https://ieu.greenclimate.fund/sites/default/files/document/220510-egm-behavioural-change-approach-paper-top_0.pdf.

Booth, Samantha, and others (2022c). Evidence review on behavioural science interventions in development and environmental fields in developing countries: Protocol (forthcoming). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund.

Appendix 3. CRITICAL APPRAISAL TOOL

The data extraction tool can be accessed in the approach paper for the evidence review and the SR protocol:

Booth, Samantha, and others (2022a). Evidence review on behavioural change in developing countries: Approach paper (February). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund. Available at https://ieu.greenclimate.fund/sites/default/files/document/220510-egm-behavioural-change-approach-paper-top_0.pdf.

Booth, Samantha, and others (2022c). Evidence review on behavioural science interventions in development and environmental fields in developing countries: Protocol (forthcoming). Songdo, South Korea: Independent Evaluation Unit, Green Climate Fund.

Appendix 4. GRADING OF RECOMMENDATIONS ASSESSMENT, DEVELOPMENT AND EVALUATION (GRADE) EVIDENCE PROFILE OF META-ANALYSIS

	QUALITY ASSESSMENT						GRADE RESULT
Intervention category (outcomes)	No. of studies (design)	Limitations	Inconsistency	Indirectness	Imprecision	Pooled effect	Quality
Feedback							
Behavioural outcomes: water and electricity consumption	5 (3 RCTs)	No serious limitation	No serious inconsistency	No serious indirectness	No serious imprecision	0.26 (0.13, 0.39)	□□□□ Moderate
Feedback and micro-incentives							
Behavioural outcomes: water and electricity consumption	2 (2 RCTs)	Serious limitation	No serious inconsistency	No serious indirectness	Serious imprecision	0.09 (-0.10,0.29)	□□□□ Very low
Feedback and social benchmarking							
Behavioural outcomes: water and electricity consumption	2 (2 RCTs)	Serious limitation	No serious inconsistency	No serious indirectness	Serious imprecision	0.03 (-0.06,0.13)	□□□□ Very low
Feedback social benchmarking and micro-incentives							
Behavioural outcomes: water and electricity consumption	2 (2 RCTs)	Serious limitation	No serious inconsistency	No serious indirectness	Serious imprecision	0.10 (-0.10,0.29)	□□□□ Very low
Feedback and goal setting							
Behavioural outcomes: electricity consumption	3 (3 RCTs)	Serious limitation	No serious inconsistency	No serious indirectness	Serious imprecision	0.11 (-0.10,0.33)	□□□□ Very low
Reminders							
Knowledge	4 (3 RCTs)	Serious limitation	Very serious inconsistency	Serious indirectness	Very serious imprecision	0.87 (0.34,1.41)	□□□□ Very low

	QUALITY ASSESSMENT						GRADE RESULT
Reminders							
Development results: crop yield	3 (3 RCTs)	Serious limitation	No serious inconsistency	No serious indirectness	Serious imprecision	0.21 (-0.03,0.45)	□□□□ Very low
Reminders							
Development results: improved income and livelihoods	3 (3 RCTs)	Serious limitation	Very serious inconsistency	Very serious indirectness	Very serious imprecision	0.44 (-0.21,1.09)	□□□□ Very low
Reminders							
Development results: improved health	2 (1 RCTs)	Very serious limitation	Very serious inconsistency	Very serious indirectness	Very serious imprecision	0.41 (-0.45,1.28)	□□□□ Very low
Reminders							
Development results: behavioural outcomes: forest use and electricity consumption	2 (2 RCTs)	Very serious limitation	Very serious inconsistency	No serious indirectness	Very serious imprecision	0.41 (-0.07,0.88)	□□□□ Very low
Goal setting							
Behavioural outcomes: electricity consumption	3 (2 RCTs)	Serious limitation	No serious inconsistency	No serious indirectness	Very serious imprecision	0.16 (-0.11,0.42)	□□□□ Very low
Goal setting and commitment devices							
Behavioural outcomes: savings behaviour	2 (2 RCTs)	Serious limitation	No serious inconsistency	No serious indirectness	Serious imprecision	0.03 (-0.06,0.12)	□□□□ Very low

Appendix 5. NARRATIVE DESCRIPTION OF FEEDBACK INTERVENTIONS

In an intervention aimed at reducing water consumption in residential households in Singapore with residents aged 18 years and above, a treatment group from the Goette and others (2019) study received a feedback message on their water consumption for the last measurement period, and an efficient benchmark. It was found that high baseline households respond more to the treatment, by saving much more water than the low baseline households. Households with a baseline higher by 1 litre reduced water use by around 0.066 litres per household (SMD: 0.0961, CI: -0.1089-0.3011). The study found that there is almost no treatment effect with low baseline households, as lower baseline households seem to increase their water use after treatment.

Brick and others (2017) consider feedback as an intervention but combine it with framing devices to target domestic water users living in free-standing houses in Cape Town, South Africa with access to an uncontrolled water supply that is metered by a credit meter to influence water conservation. The intervention entailed firstly the provision of a graphical breakdown of the bill and complemented it with a 1-page tip sheet which provided information on ways to reduce usage. The framing device intervention component was then delivered differently to two treatment groups. The financial gain treatment replicated the visual from the tariff graph treatment and additionally provided information around the potential financial savings (gain) from moving to a lower tariff block. The loss treatment replicated the information from the gain framing, but framed it as a financial loss (provided information around the financial dissaving from not moving into a lower tariff block). In line with Goette and others (2019), though in a different context, the financial gain messages treatment reduced water consumption by 206 litres per household per month on average (SMD: 0.0094, CI: -0.004 to 0.0228) and in the financial loss messages treatment reduced water consumption by 181 litres per household per month on average (SMD: 0.0076, CI: -0.4173 to 0.4326) during the intervention period.

In Nicaragua, the Barido and others (2018) study combined feedback, reminders and micro-incentives behavioural interventions to motivate households and small enterprises to reduce electricity consumption. Feedback was designed as a monthly report that included the user's electricity consumption. Participants were asked to set an electricity consumption goal for the coming month and were required to text this goal to a cloud server. The cloud server later sent an SMS in case the threshold was crossed. As a micro-incentive, user-tailored energy information, a real-time alert and a USD 6 flexible demand monthly payment was offered to participants in the intervention. Controlling for both seasonal consumption variation and federal holidays (e.g. Independence Day), with each participant in the treatment and control group being compared with itself in the previous year for every month during the intervention period, credible post-intervention month-by-month changes in mean and SD differences for the treatment group were 5.2 and 36.7 kWh per month respectively, and 36.7 and 31.5 kWh/month for the control group (SMD: -1.2252, CI: -1.7767 to 0.6737).

In Colombia, a Torres and Carlsson (2016) evaluation combined feedback with two other interventions, namely social norms and defaults. Treatment group participant households received personalized consumption reports, including a message appealing to both descriptive and injunctive norms (each household is compared to the mean and households joining the most efficient group in the current month). The information contained in the reports was based on the billed water consumption of the corresponding month. For the default component or opting-out component, households are given the option to stop receiving consumption feedback. Through an analysis

assuming no spill overs from the targeted group to the untargeted group, the study also reports evidence of intervention effects as the average household participating in the experiment reduced water use by 13 per cent (SMD: 0.1152, CI: -0.0012 to 0.2317) and 6.3 per cent (SMD: 0.0505, CI: -0.0658 to 0.1669) in the first 6 and 11 months after the start of the experiment, respectively. It is important to note that the study also finds evidence of spill-over effects: households that were not targeted by the campaign reduced water use by 5.8 per cent in the first 6 months following the intervention.

Thondhlana and Kua (2016) studied the impact of a combination of interventions, namely feedback, reminders and rules of thumb on electricity energy consumption among households in South Africa. Reminders included stickers while rules of thumb were described as clearly defined energy saving tips. Feedback was based on energy-saving performance among intervention households. After four months of the intervention, the study reported a significant reduction (-24.5 KWh) in electricity consumption between April and August among intervention groups. However, in the control group there was no significant difference (-4.81 KWh) in electricity consumption between the same time points.

Lastly, the Ruiz-Tagle and Schueftan (2021) study in Chile aimed at reducing pollution emission levels through feedback, amongst households owning a double combustion stove as their only source of heating. Feedback was provided to wood stove users through a visualization of low or high emissions levels, depending on how they chose to set their stove's damper settings. The information sign aligns to the wood stove's damper lever, thus providing real-time feedback on the wood stove's emissions at each damper setting. At the end of the intervention (after 1 month) the study found evidence of effects across emission-related measures. Firstly, average emissions of fine particulate matter (PM2.5) for those in the treatment group declined by 10.8 per cent after installation of the information sign following the intervention, and the results are statistically significant. Though this is 12.6 per cent lower than those for the control group, the effect is statistically insignificant. Second, results show a decline in imputed wood stove emissions by a coefficient of -1.844 (SMD: 0.3824, CI: 0.2727 to 1.0376). Thirdly, the probability of shifting the damper towards less polluting settings increased by 0.603 (SMD:1.6168, CI: 0.8768 to 2.3569). The study noted that since the intervention was complemented by the visit of a field assistant that explained the sign and provided the informational flyer (fridge magnet), it is possible that part of the effect may be driven by these visits acting as a more salient reminder for the treatment group than for the control group, thus modifying the effect of the information sign alone.

Appendix 6. NARRATIVE DESCRIPTION OF REMINDER INTERVENTIONS

Shah and others (2020) explored the effectiveness of a simplified account statement combined with SMS reminders, through interventions in improving Mexican participant contributions towards retirement savings compared to a standard account statement and no SMS reminders. The simplified account statement was designed to be visually appealing and readable to the participant. A firm mailed simplified or standard account statements in either paper or electronic form to the participant. In addition to the simplified statement, participants received SMS reminders with various kinds of framing encouraging participants to make a voluntary contribution. The study showed that sending simplified account statements and family security SMS reminders (framed as 'act today to improve your family's future') significantly led to an increase in savings when compared to controls at 2 months post-intervention (SMD: 0.0223, CI: -0.0012 to 0.0459) at the 10 per cent significance level compared to controls.

In order to supplement a face-to-face training on adoption of recommended practices and other agricultural outcomes, Dzanku and others (2021) tested the impact of reminders among farmers in Mali. The hypothesis was that the mobile phone voice SMS reminders intervention would improve the adoption of recommended practices and other outcomes. The study reported that reminders increased uptake of the recommended storage technology by between 11 to 18 per cent, representing an increase of 28 per cent to 42 per cent, when compared to the control group mean uptake rate (SMD: 0.2248, CI: 0.1877 to 0.6372). The study also finds no significant improvement in crop sales as measured by gross revenue (in USD) from the cereals traded (SMD: 0.0119, CI: -0.3992 to 0.4231).

To improve infectious waste management among health care workers in Pakistan, Kumar and others (2016) explored the effectiveness of a combination of training and reminders. It was hypothesized that the intervention would help sustain good health through waste management practices in teaching hospitals in Pakistan. After three months of the intervention, there was a significant statistical change in mean score of attitudes regarding infectious waste management before 27.38 (7.63) and after 34.12 (4.17) the intervention among treated health care workers (SMD: 1.1671, 95 per cent CI: 0.8695 to 1.4647). Further, the study reported a significant change in mean practices before (11.26, 4.04) and after 14.81 (2.50) intervention among treated health staff (SMD: 0.8682, CI: 0.5804 to 1.156). However, there was no significant difference in mean knowledge or practices among controls.

Using a cluster-randomized trial, Manaseki-Holland and others (2021) studied the impact of a complex intervention that included community campaigns and reminder visits at 6-month and 32-month follow-up. On evaluation, the study reported a 4.4-fold increase in the practice of the five key behaviours, such as handwashing with soap and water before preparing food, the washing of pots and utensils and drying them before use, and handwashing before food preparation, in the intervention group at 6-month follow-up (incidence rate ratio (IRR) = 4.46 (CI: 3.63 to 5.47, p-value = 0.001) when compared to controls. This significant effect was again observed at the 32-month assessment (IRR = 1.17, CI: 1.07 to 1.28, p-value = 0.002).

Appendix 7. NARRATIVE DESCRIPTION OF GOAL-SETTING INTERVENTIONS

Ashraf and others (2005) employed an RCT to determine the impact of a commitment savings device on household savings practices. The commitment savings device included a bank account and a self-set goal, and the intention was to restrict access until the desired goal was achieved. Generally, the study reported no significant impact on saving discipline (SMD: -0.1108, CI: -0.0075 to 0.2292), saving habit (SMD: 0.0439, CI: 0.0744 to 0.1622), and on saving practice in times of little cash (SMD: 0.1066, CI: - 0.0117 to 0.225) when compared to controls. Further, treated participants reported feelings of regret about their spending practices and wished for a more disciplined saving habit (SMD: 0.3384).

In an effort to fight poverty in South Africa, Dalla and others (2021) explored a programme intended to stimulate participants to develop a habit of saving in the form of a "Goal Card". The participants needed to identify a savings goal and commit to consistently saving a certain amount. The investigators wanted to know whether encouraging clients to set firm savings commitments was effective in increasing their savings. The study demonstrated a significant increase in the savings balances of participants in the goal card programme based on fortnightly collected data (SMD: 0.6394, CI: 0.4081 to 0.8706) when compared to controls.

Grohmann and others (2020) combined goal setting and planning prompts in order to determine the savings index among Ugandan small business owners. The savings index was described as follows: whether someone has any formal or informal savings; the total savings amount; the savings amount that is specifically for the business; the frequency of saving; and the savings toward a specific savings goal. Using an RCT, the investigators examined the impact of setting a savings goal and planning to achieve the desired goal on savings. The authors reported that the intervention did not influence the savings index among treated participants (SMD: 0.0621, CI: -0.1209 to 0.2451) when compared to controls.

Kast and others (2018) conducted a study that combined goal setting and micro-incentives among micro-credit clients in Chile, and examined the impact of these interventions on monthly savings balances. All participants in all groups were offered a savings account. Using an intention to treat analysis, the study reported an increase in the average monthly savings balances in the intervention group (SMD: 0.0565, CI: -0.049 to 0.1621) compared to controls. Evidence points to treated participants being encouraged to make frequent deposits and consequently savings balances increased.

Liu and others (2021) set out to examine the spill-over effects of goal setting and micro-incentives from electricity saving to water saving. Using an RCT, households in Singapore were either assigned a self-set electricity saving goal or assigned a 10 per cent goal or a 10 per cent goal in addition to micro-incentives. In terms of electricity consumption, households that were assigned a 10 per cent electricity saving goal in addition to micro-incentives, showed a significant reduction in electricity consumption (SMD: 0.2003, CI: -0.2565 to 0.65) compared to controls. Reports on the spill-over effects to water consumption indicated that self-set electricity saving intervention led to 15.1 per cent reduction in water consumption among treated households (SMD: 0.3076, CI: -0.1477 to 0.763). However, there was no spill over effect onto water consumption among households that were assigned the 10 per cent electricity savings goal only (SMD: 0.1901, CI: -0.2636 to 0.6438) or even an additional micro-incentive (SMD: 0.1433, CI: -0.3129 to 0.5996).

Appendix 8. NARRATIVE DESCRIPTION OF SALIENCE (EXPERIENCE DESIGN) INTERVENTIONS

Evaluating a different form of outcome, the uptake of energy-saving compact fluorescent lamp light bulbs (CFLs) in Kenya, the evaluation made by Figueroa and others (2019) combined salience (experience design) and micro-incentives in which participants received flyers with a graphic illustration of how a large number of incandescent bulbs are equal to one CFL. Participants were also informed about the economic savings associated with CFLs as compared to incandescent bulbs. Additionally, participants received a coupon for use in local stores to receive a discount on the price of a CFL bulb, and only applicable for CFL purchases. At endline follow-up immediately after the intervention, compared to the condition without intervention, combining these two interventions resulted in nearly ten times higher uptake rates of energy-saving CFL light bulbs (SMD: 0.4389), but this diminished at 2-year follow-up (SMD: 0.0958). The study posits that the impact of the interventions after 2 years was minimized by other trends, including changes in the overall lighting market in Nairobi, successful CFL rollout programmes, or a general spread of awareness of CFLs among the Kenyan population.

Tidwell and others (2020) combined salience (experience design) and public commitments to promote handwashing amongst school children. These took the form of presentation of interactive stories, games and songs, the use of animated characters representing occasions for handwashing with soap, the use of visual demonstrations to communicate the presence of germs and how handwashing with soap removes them in comparison to only water, the use of daily diaries by students to record their handwashing behaviour and making a public commitment as a group to handwashing with soap. Results indicate the children in the treatment reported handwashing with soap on key occasions (35.2 per cent) more than those in the control group in 20.1 per cent of the observed occasions (risk ratio: 1.77, CI: 1.22 to 2.58, p = 0.003). This programme was delivered in schools, and the impact was, accordingly, higher in schools than in the home. However, the study finds that separating the key occasions provided no evidence that handwashing with soap after defecation was higher in the treatment group than the control group (risk ratio: 1.18, CI: 0.88 to 1.57, p-value = 0.265), but did indicate strong evidence that handwashing with soap before eating was greater in the treatment than in the control group (RR: 2.68, CI: 1.43 to 5.03, p-value = 0.002).

The Yamin and others (2020) study focuses on the impact of salience (experience design), salience (communication) and social benchmarking and reminders to reduce fuel consumption amongst truck drivers in Colombia. Drivers received a small business card with their own consumption for the last month compared to the company's average and also watched a video in which a driver asks those driving fast to reduce their speed. Additionally, SMS messages were sent to drivers' mobile phones at 4, 7 and 9 weeks after the video sessions, that included each driver's consumption level of fuel. The combined interventions led to a decrease in average monthly fuel consumption (or increase in average distance travelled with one US gallon of fuel) at 1-month post-test (SMD: -0.3882, CI: -0.0949 to 0.8714) but decreased by a lesser magnitude at 2 months post-test, (SMD: - 0.2804, CI: -0.7616 to 0.2007). The magnitude increased again in the third month (SMD: - 0.4202).

Appendix 9. NARRATIVE DESCRIPTION OF SALIENCE (COMMUNICATION) INTERVENTIONS

In the Philippines, the Fiorillo and others (2014) study aimed at improving individual savings amongst poor bank clients of a bank, and combined salience (communication) goal setting, commitment devices and reminders. It involved adjusting the account opening process (using a new account opening form, a printed savings plan, a savings calendar and text message reminders) to help individuals opening the account set a savings goal, generate a feeling of commitment and personalize the experience. Participants were also sent reminders. Those in the treatment group had balances that were 37 per cent higher than those in the control group after 8 weeks of account opening (SMD = 0.6598). This result was statistically significant.

One of the treatment arms from the Miranda and others (2020) study combined goal setting, salience (communication) and planning prompts to reduce the water consumption of households in Costa Rica. It involved delivering a postcard with a household's water bill that outlined a benchmark of average water consumption, and included prompts for households to identify a consumption reduction goal and then choose actions from a list to reduce consumption. The intervention was found to be effective. It reduced the household water consumption of treatment households by 4.8 per cent in comparison to the control group (SMD = 0.0101). This finding was statistically significant at the 5 per cent level ($p < 0.05$). However, these effects persisted for 4 months and then disappeared.

The Wang and others (2021) study combines the effects of elements of commitment devices, salience (communication), micro-incentives and feedback on saving behaviour in an intervention targeted to benefit AIDS-affected children from 48 primary schools in the southwest districts of Uganda. The interventions first included workshops that focused on asset building, microenterprise development and future planning where caregiving families were invited and encouraged to participate in these workshops together with their children. Second, a Child Development Account from which matched savings could only be used for either secondary education or microenterprise development formed part of commitment device. Additionally, children would receive an equivalency of the amount their caregivers saved. Children also received a bank statement from the research team periodically, which provided feedback on amounts of savings and matches. The study found evidence of intervention effect through a decrease in the poverty (poverty incidence rates) noticeable in year 1 to year 3 after the intervention, but not in year 4, when the effects became insignificant. Whilst the study does not explore the reason for this decline, the authors note it as an important finding that warrants further investigation to understand the longer-term impacts of the intervention or the evolution of the intervention.

Wang and others (2018) used a controlled trial in Uganda to examine the impact of a Bridges and Bridges Plus programme on economic outcomes among low-income AIDS-affected children in 48 primary schools in a rural region of Uganda. The programme consisted of two behavioural interventions referred to as Bridges and Bridges Plus. Bridges and Bridges Plus consisted of micro-incentives, a child development account and salience communication. The only difference between the two interventions was the level of financial micro-incentives the participants received through their efforts to save money. To further explain the difference between Bridges and Bridges Plus, participants in the Bridges intervention received a 1:1 match rate as a micro-incentive, that is they received an equivalent of USD 1 for each USD 1 they deposited into their development accounts, while participants in the Bridges Plus intervention received a 1:2 match rate, meaning that for each USD 1 equivalent they deposited into their own account, they received an equivalent of USD 2. There was a significant increase in the log amount of savings at 24 months among the Bridges-

treated participants (SMD: 0.1619, CI: 0.0291 to 0.2947) at 0.1 per cent significance level when compared to controls. Further, the study showed a significant increase in log amount of saving among Bridges Plus-treated participants (SMD: 0.3148, CI: 0.1893 to 0.4404).

The Young (2017) study engaged Malawian smallholder farmers in a training on various subjects including agri-ecology, climate change, and soil health, among others. Investigators wanted to understand if including participatory drama in the training would influence social interaction, attitudes and efficacy so that farmers can understand, share and sustain knowledge. When compared to controls, the study did not find significant differences in knowledge of inter-relationships between sustainability issues with regard to climate change and human health (SMD: -0.1915, CI: -0.4103 to 0.0273). Further, there was a minimal increase in reports of communication among treated participants and controls (SMD: 0.1178, CI: -0.2695 to 0.5051). Increases in information sharing was reported in control participants (SMD: -1.0185, CI: -1.3923 to -0.6446). Narrative and participatory drama intervention influenced a minimal increase in the saving of resources (SMD: 0.3449, CI: 0.0285 to 0.0285) when compared to controls.

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