A review of the literature

Learning lessons for education from the use of results-based financing (RBF) in health

Social Finance Ltd, Percept Actuaries & Consultants, JET Education Services

3 December 2021

SOCIAL FINANCE

PERCEPT
HEALTH · SOLVE

JET EDUCATION SERVICES
THINK EDUCATION. THINK JET.

WORLD BANK GROUP

Acknowledgements section

This study was funded by the Results in Education for All Children (REACH) trust fund at the World Bank Group. REACH is supported by the governments of Germany, Norway, and the United States. For more information, please visit: www.worldbank.org/reach

Authors

1.1 Rose Tuyeni Peter, Economist - Percept

Rose Tuyeni Peter is an economist with experience in applied research, specifically in health, economic evaluation, financial inclusions, and financial sector development. She has completed an MSc in Health Economics and Decision Science at University College, London. Prior to extending her economics work into health, Rose had worked as a Senior Research Analyst at the Centre for Financial Regulation and Inclusion (Cenfri) and completed a master's degree in economics at the University of Cape Town.

1.2 Helen Redpath, Consultant - Percept

Helen works as an economist on competition/ antitrust, innovation, policy and regulation projects, with her work spanning all sectors. She has over a decade of experience in these areas, having worked in an investment bank, a local economics firm, and a global economics firm, as well as with a range of other firms. Helen currently works as an independent economist for both local and global firms, including Percept. Helen holds a MSc in Economics cum laude from her studies at the University of Bologna (Italy), and completes ongoing professional training at the Barcelona School of Economics. She writes regularly and has published in local and international journals. She is passionate about making the economy more accessible and helping businesses thrive.

1.3 Louise Savell, Director – Social Finance Ltd.

Louise Savell is co-founder of Social Finance and jointly leads their international development team where she advises governments, philanthropies and service providers on the design and delivery of social development programmes. Louise is passionate about driving social impact through rigorous analysis, efficient structures and effective cross-sector partnerships. She has particular expertise in outcomes-based approaches and social investment structures. She co-developed the Impact Bond approach in 2008. Louise has led RBF work across a range of sectors and geographies including Sleeping Sickness in Uganda; Education in Ghana and Liberia; Early Childhood Development in China, South Africa and Cameroon; Energy Access in India; and Street Children in Ethiopia. She has also led the development of RBF programs in the UK in the fields of homelessness, youth employment and health.

1.4 Anja Smith, Consultant – Percept

Dr Anja Smith is a development economist and part-time academic. After spending eight years of her career consulting in the area of access to financial services and financial inclusion, she pursued a PhD in economics with a focus on health systems on a full-time basis. Her PhD and postdoctorate focused on topics related to the financing, delivery and quality of healthcare in South Africa. Anja works as an

independent economic expert (also for Percept) and a researcher at Research on Socioeconomic Policy (RESEP) at the Economics Department of Stellenbosch University.

1.5 Nick Taylor, Senior Research Fellow – JET Education Services

Following 10 years of teaching high school maths and science (1975-1984), Nick was appointed Subject Advisor for Mathematics for the Soweto Region of the Dept of Education and Training (1985-1988). He then joined the Education Policy Unit at Wits University as Policy Researcher, where he was involved in exploring education policies for a democratic South Africa. This experience was consolidated over the years through a number of initiatives, including leadership of the President's Education Initiative (1999), a number of large school improvement programmes funded by DFID, DANDIA, USAID and South Africa's Business TRUST, and the National School Effectiveness Study supported by the Royal Netherlands Embassy. Between May 2012 and April 2015 Nick was seconded to the Ministry of Basic Education, where he established and ran the National Education Evaluation and Development Unit (NEEDU), a national school evaluation unit. He led the Initial Teacher Education Research Project (ITERP), to investigate the quality of pre-service teacher education. ITERP led to the Primary Teacher Evaluation Project (PrimTEd), directed towards developing standards in mathematics and literacy for primary school teachers in training, led by the Dept of Higher Education and Training.

1.6 Zaahedah Vally, Researcher – JET Education Services

Zaahedah Vally is a researcher at JET Education Services based in South Africa, working across research and implementation projects. She received her Bachelor's degree in International Relations with the University of South Africa (UNISA) and a Postgraduate Diploma in Management (Monitoring and Evaluation) with the University of Witwatersrand. Currently completing her Master's degree in Management (Development and Economics), Zaahedah is passionate about driving socio-economic change in South Africa by improving the quality of education in the country. Her Master's report will focus on the readiness for the adoption of outcomes-based contracting in South Africa.

1.7 Jodi Wishnia, Public Health Consultant - Percept

Jodi is a public health specialist and executive at Percept. She has an undergraduate degree in Occupational Therapy (BSc) and a master's degree in Public Health. Her PhD focused on public financial management in public health systems. She has published in the BMJ Global Health journal and contributes annually to the South African District Health Barometer on the relationship between South Africa's health spending and health system performance.

Research consortium

1.8 JET Education Services

JET Education Services (JET) is an independent non-governmental organisation founded in 1992 which works with government, the private sector, international development agencies and education institutions to improve the quality of education and the relationship between education, skills development and the world of work. JET's focus areas in the education sectorare education research and planning; monitoring and evaluation of education and training programmes; school and district improvement; and Technical and Vocational Education and Training (TVET) College improvement and youth livelihoods.

JET has over 18 years of experience of conducting large-scale education projects and has been successfully involved in education research, evaluation and assessment projects for both government and other sectors. JET's range of work includes implementing hands-on school improvement projects, teacher training, school turnaround programmes, managing provincial systemic assessment contracts and working with both national and provincial governments on large-scale school surveys and research projects. JET also has experience in working on systemic assessments and has proven ability to manage evaluation projects of this nature effectively.

1.9 Percept

Percept is a transdisciplinary health and finance sector consulting firm. Percept brings with it deep expertise in the public and private health sectors, in South Africa and several other African countries (Uganda, Kenya, Malawi,etc.).

Percept has extensive experience in the area of linking financing to outcomes gained through various projects on alternative reimbursement approaches and the broader topic of health financing in the health sector. Results-based financing, or payment for outcomes through alternative financing arrangements, was most recently explored in a project for the South African National Treasury on a potential impact bond associated with MomConnect, the public sector's largest telehealth platform used for communication and messaging to pregnant women and mothers.

1.10 Social Finance

Social Finance is a UK-based not-for-profit consultancy at the forefront of designing and delivering results-based financing (RBF) structures to improve social outcomes for vulnerable populations. With a specialist team of 100 staff, we are a globally-recognized advisor and impact investment intermediary that, since our inception in 2007, has pioneered, designed and delivered a series of innovative results-based financing programs around the world and across a range of countries and sectors, including health and education.

Social Finance pioneered the concept of impact bonds, launching the first ever example in 2010, an innovation that has now been taken up in multiple countries and in the field of international

development, with over 200 examples now launched globally. This has catalysed a new field of outcomes-based impact investment. Building on this, the organisation has led the way in integrating results-based finance with other forms of innovative and traditional finance.

Contents

Acronyms and abbreviations

Acronym	Definition
ARV	Antiretroviral
CAL	Computer-assisted Learning
CCT	Conditional Cash Transfers
CRBF	Community Performance-Based Financing
CPD	Continuous Professional Development
DfID	Department for International Development
ECE/D	Early Childhood Educare/Development
FAS	Foundation Assisted Schools
HICs	High Income Countries
HIV	Human Immunodeficiency Virus
ITE	IT Enabler
LMICs	Lower- and Middle Income-Countries
M&E	Monitoring and Evaluation
MNCH	Maternal, Newborn and Child Health
NGO	Non-Governmental Organisation
NULP	Northern Uganda Literacy Intervention
OECD	Organisation for Economic Co-operation and Development
PBA	Performance Based Aid
PBC	Performance Based Contract
RBF	Performance Based Financing
PBT	Performance Based Transfer
PFP/P4P	Pay for Performance
PICO	Population, Intervention, Comparison and Outcomes
PISA	Programme for International Student Assessment

PMTCT	Prevention of Mother to Child Transmission
PROGRESA	Program de Educacion, Salud y Alimentacion
RBA	Results-Based Aid
RBF	Results-Based Financing
RBF4MNH	Results-Based Financing for Maternal and Newborn Health
REACH	Results in Education for All Children
RLL	Read-Learn-Lead
SDG	Sustainable Development Goal
SNED	Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales Subvencionados
SRP	School Readiness Programme
WHO	World Health Organization

Key Terms

Term	**Definition**
Antenatal	During or relating to pregnancy
Apgar test	A test used to evaluate a newborn's health, measuring heart rate, respiratory effort, muscle tone, skin colour and reflex irritability
Decentralisation	The transfer of control of an activity or organisation to several local offices or authorities rather than one single one
Disbursement	The payment of money from a fund
Hermetically	In a way that is insulated or protected from outside influences
Meta-analysis	Examination of data from a number of independent studies of the same subject, in order to determine overall trends
Pedagogical	Relating to teaching
Pedagogy	The method and practice of teaching, especially as an academic subject or theoretical concept

1. Introduction

Our research team, comprised of JET Education Services, Percept, and Social Finance, has been contracted by the World Bank to summarise what can be learned from results-based financing as applied in the health sector in order to inform RBF in the education sector, particularly in low- and middle-income countries (LMICs) and with attention to improving primary health and basic education (pre-primary, primary, and secondary education).

Results-based financing (RBF) is a catch-all term which describes a program or intervention which rewards individuals or organisations after they have achieved an agreed upon set of results and this achievement has been verified based on a pre-determined set of targets.[1,2] Therefore, RBF ensures that development funding is directly and verifiably linked to pre-agreed targets or results and funding is only disbursed once these targets have been reached.[2] The goal of RBF is to deliver on development outcomes, improve mechanisms of accountability, and promote efficiency and innovation in the way these outcomes are achieved.[2] The emphasis on providing quality education in the Sustainable Development Goals has driven a global agenda focused on improving education in LMICs. Of the 17 SDGs[3], Goal 4 refers to education, and is stated as:

Ensure inclusive and quality education for all and promote lifelong learning.

Goal 4 has ten targets, the first of which concerns basic education (primary and secondary schooling):

By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and Goal-4 effective learning outcomes.

Much progress has been made in moving towards the access related aspect of this target, namely, universal access to primary schooling (greater numbers in school). However, with respect to both equity (boys still outnumber girls in many countries) and quality (learning outcomes) much remains to be done.[4] Above all, raising learning outcomes has come to constitute a major focus throughout, not only in LMICs, but in many high income countries (HICs) too, including the US. Despite large increases in enrolment globally, the World Bank (2019) has introduced the concept of "learning poverty", which highlights the learning crisis, especially in LMICs, where almost 53% of learners cannot read proficiently by the age of 10.[5] According to the World Bank (2021), this crisis is likely to worsen with school closures due to COVID-19, with the learning poverty rate increasing to 63% in LMICs.[6] Because literacy is a gateway to education, the inability to read means that many other areas of learning are negatively impacted.[7] There is widespread agreement that education interventions are more impactful the earlier they are introduced.[8] Furthermore, other critical foundational skills – numeracy, digital and transferable skills like problem-solving and critical-thinking are not being developed currently in education systems across the world, leaving millions with poor employment prospects.[9]

A vast number of authors refer to the relationship between teacher quality and student learning outcomes[4,10–12] and conclude that high teacher quality has a positive impact on learning outcomes, especially in LMICs. The dimensions of teacher quality differ across studies and Bold et al. (2017) used teacher content knowledge, pedagogical knowledge, pedagogical skills, and attendance as measures of

teacher quality, while Glewwe et al. (2011) use teacher qualifications as a measure of teacher quality.[10,11] The latter definition has clear drawbacks in that it does not assess actual skill, relying purely on the attainment of qualifications as proof of quality.

The health sector provides an important knowledge base to draw on, both from the perspective of using RBF and in terms of the challenges it faces in ensuring quality service provision, particularly in LMICs. Poor quality of healthcare has been a key driver of poor health outcomes in LMICs.[13] Similar to education, quality is often overlooked in favour of access to care, which remains at the top of national agendas. Indicative of this is that nine million individuals die annually as a result of inadequate quality of care despite having treatable conditions and accessible health facilities.[13] Furthermore, approximately 60% of deaths occur among those who did access healthcare services, highlighting the need for quality reforms.[13]

Beyond the access versus quality comparison, several other parallels exist between the health and education sectors, and we expand on this further below. As a result, lessons from interventions in the health sector can be valuable for the education sector. This project will leverage the lessons learned from RBF efforts in health systems. RBF has been used in the health sector to try to improve quality of care. Specifically, this literature review focuses on the use of RBF for improved health worker performance. The lessons from these interventions provide ample learnings for the education sector, as they consider the role RBF might play in improving education quality and outcomes.

The remainder of this document is structured as follows:

- Section 2 provides an overview of our key conclusions and recommendations;
- Section 3 provides context for the literature review;
- Section 4 describes our approach to identifying relevant papers and our framework for comparing the education and health literature;
- Section 5 presents the main findings from our review of health literature;
- Section 6 adds useful further insights from the education literature; and
- Section 7 discusses the main findings and emerging conclusions.

2. Overview of key conclusions and recommendations

2.1 Background

- As significant progress has been made over the last 10 years, the emphasis in education in LMICs has shifted from ensuring school access to ensuring that children are in school and learning.
- Poor quality of healthcare has also been identified as a key driver of poor health outcomes in LMICs, yet is often overlooked in favour of access to care which remains at the top of national agendas.
- Beyond this access vs. quality comparison several other parallels exist between health and education which we explore below. As a result, we are comfortable that lessons from Results Based Finance (RBF) interventions in the health sector can be valuable for education.
- An extensive body of literature exists around the use of RBF in health, although recent meta-analyses highlight that evidence is often still fairly weak, not because RBF doesn't work, but because studies are often not designed to elicit accurately attributable results nor to identify the exact mechanism by which RBF works.
- Given this context, and the global emphasis around improving education quality (sometimes referred to as reducing 'learning poverty'), our literature review has particularly focused on examples where RBF has been used in the health sector to improve quality of care.

2.2 Comparing the education and health sectors in the context of RBF

Similarities

- Health and education are both considered public goods and are often provided by the State, particularly at a primary healthcare / basic education level. Interventions in both sectors become more expensive as they become more complex (e.g. hospitalisation and tertiary / higher education).[14]
- Socioeconomic factors (social determinants) are major determinants of both health and education outcomes. Limiting access based on ability to pay has significant impacts on population health and economic growth.[14]
- Although health and education inputs – adequate and appropriate human and physical resources – are important for delivering positive health and education outcomes, social determinants also play a role.[14]
- Both health and education sectors tend to have strong unions or professional associations which protect the interests of healthcare workers and educators, but can be hostile towards the use of

incentives to control the cost or quality of how these services are provided, and any new accountability mechanisms to monitor the impact of interventions.[14]

- Common incentive problems in health and education that RBF might address include:
    - Supply limitations for key infrastructure and / or consumable resources (structural quality);
    - Insufficient or inappropriate performance incentives for personnel (process quality); and
    - Structures and / or cultures that discourage or prohibit adaptation and innovation that could improve outcomes at a facility or individual level (autonomy).

Differences

- Although there are many parallels between the education and health sectors, there are also key differences that may influence relative effect sizes and appropriate incentive mechanisms.
- Education is a more predictable service to run than health – the number of learners each year is more predictable than, particularly acute healthcare needs (as the Covid-19 pandemic has amply demonstrated).[14] Engagement with healthcare also tends to be more episodic throughout individuals' lifespan as opposed to education which tends to be discrete and has a defined start and end point.[14] Hence why, as de Moura Castro and Musgrove (2016) state, there is a need for health insurance but no market for "education insurance".
- The healthcare system has a hierarchical organisational structure, whereas education is sequential.[14] In other words, in education learners need to pass primary school to progress to secondary school and then tertiary education, in health patients are referred up or down from one level of care to another. This is sequential nature of education is because it is cumulative, whereas healthcare is not.
-
- Structural quality – infrastructure and consumable resources – is more critical to an effective health service, than an effective education service. High quality educators can compensate for many resource limitations, particularly in early grade literacy and maths education, given the autonomy and creativity to do so. It is more challenging, even for highly skilled health workers to overcome a shortage of vaccines or essential medicines which are required to treat or prevent certain diseases.
- Compared to the education sector, there is more of a reliance on donor funding in the health sector in LMICs, particularly for RBF programmes. RBF programmes are almost always donor initiated and/or supported and domestic funds are rarely devoted to these interventions.[15]
- Overprovision of health services in fee for service (FFS) models of healthcare are widely reported, particularly in the private sector.[16,17] This is likely because some healthcare providers are paid per consultation and the health services provided in each consultation can be charged in an

itemised way. This allows the cost of each input to be charged separately, and additional (potentially superfluous) services or consumables bundled together in one consultation to increase the price of care which increases the revenues of healthcare providers but make health services unaffordable for payers.[17] Therefore, it is believed that in FFS healthcare models incentivise the overprovision of health services.[16,17] However, because education services cannot be disaggregated and charged in the same itemised way and medical consultations can be, the education sector does not face the same problems caused by overprovision that the health sector does. Indeed, while much progress has been made towards achieving SDG4, the under-provision of education remains a significant problem in many LMICs. Thus, although the number of out-of-school children of primary school age in the world had declined from 99,7 million in 2000 to 59,1 million in 2018 the latter figure nevertheless represents 8% of this population fraction and has stagnated since 2008[18].

2.3 Lessons from RBF in health for education

Our review of the RBF literature in the health sector revealed six axes of programme design considerations that are important when considering effectiveness (Figure 2). The following key lessons emerged from our analysis around these themes.

1. **Understand the context:** socioeconomic determinants, the source and availability of funding (i.e. government vs donor funding), and the state of the responsible system (i.e. government interests, systems and processes) need to be considered in order to diagnose the problem, clearly articulate the nature of the problem to be addressed and the mechanism by which RBF could help. Political economic context of the health system such as the locus of decision making power and public financial management processes also influence the sustainability of RBF schemes (see section 5.1).

2. **Be clear about what is being incentivised and the expected quality improvement process:** structural and process quality are both important to driving better outcomes. Health RBF literature revealed the importance of keeping incentives clear and simple; allowing sufficient time for the incentive programme to be understood and for results to materialise; and encouraging positive feedback loops and dynamic effects by allowing facilities the autonomy to use incentive payments for both staff payments and additional facility inputs (see section 5.2).

3. **Ensure that those who are incentivised have control over the targeted actions or outcomes:** this may sometimes require incentives throughout the health or education value chain to be considered, to avoid supply constraints and ensure all relevant stakeholders are aligned in their activities and priorities (see section 5.3).

4. **Ensure strong alignment around who is incentivised, and how incentives are measured and paid:** where funds do not reach those whose actions are incentivised, success may be limited or unsustainable. Incentives at the facility level may encourage teamwork – particularly where co-workers actions are observable to others and individual bonuses are conditional on achieving institutional targets (see section 5.3).

5. **The functions of purchasing and providing services should be split:** when the purchaser and provider are split, the accountability of providers is improved because they no longer reimburse themselves, they are reimbursed by the purchaser of services. The purchaser is able to build in payment incentives, such as RBF much more simply and with more credibility, because the verification of results and monitoring of performance is separated from those who stand to gain from the incentive payments (or the non-payment of incentives). Valuable RBF schemes therefore require intentional separation of functions in order to verify achievements (see section 5.3).

6. **Consider the size, timing, and form of payments:** the size, timing, and relationship of payments to marginal effort all matter. More frequent payments increase the saliency of RBF programmes and higher and / or easier to achieve payment triggers increase the likelihood of incentives influencing behaviours. Individual incentives seem to be more influential when received as bonuses rather than basic salary payments (see section 5.4).

7. **Adjust for equity and reward absolute, rather than relative, progress:** higher incentive payments for facilities with lower resources may offset differences in initial starting points across facilities. Targeting incremental improvements as opposed to using payment thresholds, and rewarding absolute rather than relative performance, both reduces unintended consequences and increases staff motivation (see section 5.4).

8. **Enhance support and supervision to frontline staff:** several studies identified that enhanced, regular supervision and structured feedback is important to achieving improved outcomes. Detailed checklists, which sometimes also form part of assessing whether the outcomes which trigger disbursements have been achieved, present a useful and structured opportunity to assess the performance of individuals and facilities, and provide timely feedback on how to improve (see section 5.5).

9. **Enhanced supervision and financial support are reinforcing:** Although enhanced supervision and monitoring are necessary for improving service delivery, additional financing is also required. One study theorises that measurement and feedback of health worker care quality plays a primary role in improving health outcomes, and the bonus payment acting as an “accelerator” for these improvements (see section 5.5).

10. **Recognise the need for delivery autonomy and support adaptive service delivery:** RBF is most effective where facility managers have the autonomy to adapt their services to improve outcomes. However, initially such approaches may be extremely new and facility managers may need support to enable adaptive decision-making and improved human and physical resource management that ensure that bottlenecks are effectively identified and cleared (see section 5.5).

11. **Embed processes to monitor for and correct potential unintended consequences:** build in independent audit processes and, potentially, include financial penalties for inaccurate reporting of programme results. Monitor and, if necessary, adjust for demand-side or supply-side impediments, that may not be within the control of the incentivised facilities or staff (see section 5.6).

12. **Adequately fund impact and process evaluations:** well-powered, independent evaluations of programme results and ideally processes, will help to strengthen the evidence and knowledge base around what works in terms of both effective interventions and robust RBF programme design (see section 5.7).

2.4 Areas where RBF may be most valuable in education

The intractability of raising test scores points to the need to focus interventions directly on the knowledge and skills of teachers. Programs directed towards strengthening the foundation disciplines in young learners build the tools required for all subsequent learning. It is no surprise that research findings indicate strongly that the earlier in the education process teacher capacity building is instituted, the stronger the effects on learning outcomes[19]. In recognition of these conclusions, the overwhelming priority for donors and governments alike is to focus on teacher pedagogic capacity at ECE and early primary levels[20]. However, pedagogically-focused interventions require that schools are equipped for learning: that management support and monitoring services are in place and effective, that textbooks and stationery are supplied, and that communities are involved in governance (see Appendix 1). These considerations dictate that interventions aimed at improving learning outcomes are necessarily complex and may involve direct support to teachers, school-level grants and support to learners and families. Donors are funding these various elements of effective schooling, and we discuss each of them in more detail below, excluding demand-side interventions such as cash grants to families.

Understand the problem

Understanding the problem to be addressed is fundamental to program design. If the problem is teacher attendance, it is well to note that incentives to teachers to attend more frequently may be successful in the short term but are prone to gaming[21] and, we argue below, run the risk of undermining discipline and morale in the civil service (see Sections 3.1, 5 and 6). However, the only education programs which consistently improve learning are those targeted specifically at teacher capacity[20]. Two broad design categories have emerged as particularly effective: structured pedagogy and teaching at the right level (TaRL) (Appendix 1).

Understand the context

Understanding the political economy into which a project is to be introduced is key to success. This is true both at macro, state level and in the immediate environment of the project. Political rivalries, ministerial vanity projects and cabals involved in corruption may be decisive in frustrating donor intentions[22]. Indeed, government buy-in is essential even for small scale experimental initiatives, and without active government participation from the earliest stages, take-up by the state educational system is obviously impossible[22]. Civil service competence, including the use of data for monitoring all moving parts of implementation is important for government participation and eventual take up[22].

In implementing RBF projects in the health sector, studies showed that obtaining the buy in of the health workers at the facilities whose performance is being evaluated is essential to the success of interventions. To achieve this, the purpose of the intervention, how incentives are paid out, and the interventions coherence with national policy objectives needs to be explained. Where contextual factors that influence

the ability of facilities to attain targets differ across facilities, such as resourcing, then equity weights or bonuses should be applied so that incentive payments are not perceived to be unfair.

Both health and education consist of complex sets of institutions, and program success may depend on taking account of systemic considerations. Therefore, where interventions rely on the coordination of multiple stakeholders at different tiers of the system, incentives will need to be targeted at actors at each of these levels, so that they are motivated to work towards a common outcome (see section 5.3 for a more detailed discussion).

Knowledge building

Donors are showing an increased interest in understanding what works in strengthening the knowledge base around effective interventions in raising educational outcomes. At the project level, strong verification systems are needed to track implementation and make course corrections. Studies cited in the health literature show that rigorous verification processes are important for ensuring that outcomes reported are a true reflection of the outcomes obtained, and these verification procedures may present an opportunity to foster improved communication and coordination between those implementing interventions and those being incentivised.

As part of the evaluation process, it is crucial that RBF interventions be observed over a sufficient period, at least 18 months as noted in one study,[23] to truly understand their impact on outcomes otherwise the interpretation of these results may be myopic or misleading. A combination of experimental studies and qualitative observational studies is needed to reveal impact and to understand the mechanisms of change or stasis[24]. We argue above that the problem of external validity is insufficiently addressed in the field, although we quote examples of education projects taken from proof-of-concept to systemic embedding, most projects are small in scale and of short duration[20]. Under such artificial experimental conditions, no matter how big the effect sizes, conclusions can only be provisional and tentative, and may not translate to success at regional or national scale[25].

2.5 Areas for further research

RBF in health

More is needed to unpack what is described as the RBF "black box." This "black box" refers to the many studies which describe RBF interventions and provide quantitative analysis of their impact on targeted outcomes, but they provide limited descriptions of the design features of interventions which are the most important for improving these outcomes. Given the complexity of certain composite measures of quality (such as the indices described in section 4.2) more qualitative analysis may be valuable in illuminating why interventions may work in some contexts as opposed to others. Further understanding of the design features which work best also begs the question of how RBF fairs compared to other interventions which aim to strengthen health systems, such as social impact bonds and non-incentive-based interventions.

Many of the health studies reviewed mentioned little in the line of support of the RBF program, for example to help those incentivised adapt to a new way of working, help facility managers make decisions,

help incorporate feedback to adjust the program if needed, etc. Those studies that did mention this highlighted its importance, but also mentioned that a great deal of resources go toward verification of the data, lessening the resources left for this kind of support. Research into the amount and type of support needed and the appropriate resources and people for this would be beneficial. It would be useful if more studies added an explanation of this, possibly also with mention of communication with those on the ground (for example, facility managers) as to what type of support is needed at different points in time and how this informs program and study designs, both initially and iteratively.

RBF in education

The same issues noted in the previous paragraph with respect to research needs in health apply in education, only more so given that the use of RBF in education is at an earlier stage of development. The success of pedagogically-focused interventions in raising test scores opens the possibility of applying RBF funding modalities to such models. It is our contention that there seems to be no in-principle reason why such models should not be as, or more, effective as those which use non-RBF approaches to funding, an argument we take up in section 6.2.

A second research priority is to understand the conditions under which programs which have established proof of concept are taken to large-scale implementation. Since going to scale generally requires government take-up, this research task is a multi-faceted one, including the political economy into which the project is intended to be inserted and the extent to which the program is aligned with official policy, the capacity and commitment of the government bureaucracy, and the cost effectiveness of the program and the long-term budgetary implications.

Third, where non-government institutions are supported in order to increase pupil enrolment and learning outcomes, such as government subsidies to low-cost private schools in a number of African countries, success has been found to be mixed[26]. Such programs have been found to be effective under conditions of weak government accountability systems[27]. Therefore, understanding the specific conditions under which they are successful holds much promise for expanding access to and success in schooling for children from low-income homes.

3. Context for the literature review

3.1 What has already been done in education?

The Results in Education for All Children (REACH) project has undertaken a research investigation into the use of RBF modalities to improve a number of aspects of basic education.[21] This was a wide-ranging study, encompassing review of experimental and quasi-experimental studies, meta-analyses, project documents, a survey among donors and follow-up interviews. The study found some evidence for effective RBF experiments with regard to incentive programs applied to teachers, schools and government, but overall the evidence suggested that there is room for further study.

Some pertinent learnings from the REACH study include learnings under the following broad themes:

- **Teacher incentives.** Teachers play a critical role in determining the scope and quality of students' learning, yet in many countries, they lack the support, motivation, and training to do their jobs well. Teachers may not be well versed in the subjects that they teach, or don't have the pedagogical skills to effectively transmit knowledge to students. Teacher absenteeism is a further problem. As a result, many children worldwide are failing to master basic skills. The REACH report concludes that incentive schemes can motivate teachers by rewarding good performance, encouraging them to expend more effort in teaching and preparation, and to show up in the classroom. Evidence on the effectiveness of teacher incentives in raising learning is limited, due in part to the small number of interventions that have undergone rigorous evaluation. Findings regarding the relationship between these measures and learning outcomes are mixed. An evaluation of a programme which rewarded teachers with incentives based on students' test scores in rural Kenyan primary schools found that teachers increased test preparation and scores increased in subjects linked to the incentives, but scores for unrelated subjects did not increase.[28] Students' scores in treatment schools increased, however this did not last beyond the lifetime of the programme. Teacher attendance did not improve, homework assignment did not increase and the dropout rate of learners remained unchanged during the programme.[28]
- **School grants.** School grants provide the discretionary funds schools need to implement their own improvement plans. While many grants are given to schools without conditions, some countries are beginning to link funding to school performance in an attempt to incentivize schools to focus on their central objective to improve student learning. Unconditional school grants, although not an instrument of RBF, have improved enrolment and raised student completion rates but have been less successful at improving student learning. There are also models for low-cost private schools where loans are made to fund school expansion/ resource purchases with an element of loan forgiveness if the schools demonstrate improved learning outcomes during the repayment period.[29]

- Impact Bonds[1] (outcomes-based contracts in which investors pre-finance service delivery against a pay-for-results contract with government or donors) in the education sector are in their early stages, however there is increasing traction in this area particularly around early childhood education, basic education, and vocational training for employment. The development sector is moving away from traditional input- and activity-based funding models towards results-based approaches in which learner outcomes are prioritised.

The education literature search (and the focus when drawing learnings about RBF from health for education) on which the remainder of this report draws has been guided by three principles:

- A principal focus on interventions that seek ways of improving learning outcomes. Although the battle for school access has not been won universally, there have been very significant gains in school attendance around the world, particularly in low-income countries. Yet, in these countries learning success is poor, and attention has moved to targeting the improvement of what has been called epistemic access,[31] as measured by test scores. There is now a strong focus on early grade reading and mathematics, as the foundation disciplines for all subsequent learning. Access remains a secondary focus.
- Since the number of projects utilising RBF modalities (and which fit the criterion listed above) is low, we also searched for interventions which have demonstrated success in enhancing learning outcomes and/or increasing access but which used traditional funding modalities. With respect to these programs, we ask the question: what is needed, if anything, to apply RBF approaches to such interventions?
- Meta-analyses in both fields of the literature mentioned above frequently produce 'mixed results'. A likely explanation for this phenomenon is that under certain conditions success is facilitated but inhibited under different circumstances. This is why an understanding of the context in which a program is to be applied is critical to program design. The priority research question under these circumstances is to understand the differences between those initiatives where the results are positive and those in which they are negative, in order to tease out the specific factors which seem to contribute

[1] More specifically, 'Impact Bonds' are described by Government Outcomes Lab (GO Lab), the University of Oxford team which hosts the Global Knowledge Hub for outcomes based approaches,, as *"outcome-based contracts that incorporate the use of private funding from investors to cover the upfront capital required for a provider to set up and deliver a service. The service is set out to achieve measurable outcomes established by the commissioning authority (or outcome payer) and the investor is repaid only if these outcomes are achieved. Impact bonds encompass both social impact bonds and development impact bonds."*[30] Outcomes contracts and impact bonds are differentiated from other forms of RBF by tying payments to verified service user outcomes, as opposed to service provider activities or inputs.

towards a positive impact. This, in turn, requires a detailed analysis of individual initiatives in the form of case studies.

Social and Development Impact Bonds

Social and Development Impact Bonds are part of a family of outcomes-based approaches which revolve around funding that is tied to the achievement of substantive outputs and outcomes, as opposed to inputs, activities or process milestones. The ambition is that by contracting for outcomes, rather than activities, service providers will have greater scope to flex and adapt their delivery approach to respond to changes to the delivery context and / or population needs, which in turn will enable them to maximise outcomes for service users.[22]

Since the first Social Impact Bond contract was launched by Social Finance in the UK in 2009, 223 Impact Bonds have been launched of which 30 have been in education, but only 2 have completed to date.[23] Notably, the incentivised party in almost all impact bonds tends to be a non-state provider of capacity-building services and only sometimes includes incentive payments to teachers and / or health workers in the supported facilities. While there is increasing interest in the potential of such mechanisms to enable service innovation and adaptation in changing contexts, insights around their effectiveness and the underlying drivers in terms of programme design are necessarily limited by the extremely small sample.

The Educate Girls Development Impact Bond (DIB) – Rajasthan, India

The Educate Girls DIB in India was the very first DIB in education to be launched in 2015. When it concluded in July 2018, it had exceeded the target for the pre-defined outcomes of learning gains of boys and girls in grades 3-5 and enrolment of out-of-school girls.[24] The three main lessons learnt from this first DIB were: innovation is stimulated on the implementers part whilst they attempt to achieve results; if targets are not carefully measured and evaluated, incorrect payments may be made and ineffective or harmful programme adjustments might be made; and there is opportunity to further tests DIBs provided that they are contextualised for specific needs and important lessons from previous DIBs be considered.

The Impact Bond Innovation Fund – Western Cape, South Africa

The Impact Bond Innovation Fund (IBIF) was the first early childhood development SIB to be launched in South Africa and was the third impact bond globally with the government as the outcomes funder in a LMIC country. It sought to improve early childhood learning and development outcomes in the Western Cape. It specifically targeted 3-5 year-olds in two impoverished communities in the Cape metro area: Delft and Atlantis. The Western Cape Foundation for Community Work provided home-based early learning services to preschool-aged children.

The programme exceeded targets for the target of recruiting and retaining 2000 children in the programme over three years, however the early learning outcome measure targets were missed despite improvements against this measure.[25] This was judged to be mainly due to the ELOM test being administered in home settings rather than being centre-based (*ibid.*).

3.2 What has already been done in health?

An extensive body of literature exists around the use of RBF in health. That being said, recent meta-analyses highlight that the evidence is often still fairly weak, not because RBF does not work, but because studies often are not designed to carefully elicit accurately attributable results and identify the exact mechanism by which RBF works. This may be for a myriad of reasons, including that a counterfactual was not well set up, the mechanism of improvement was not carefully studied, or the starting point was not well accounted for (improvement from a good base may naturally be less than improvement from a poor base).

Since 2007, the Health Results Innovation Trust Fund (HRITF) has been supporting results-based financing (RBF) approaches in the health sector to improve maternal and child health globally.[32] As part of the evaluation of their learning portfolio, they summarized the findings of seven of their completed impact evaluations in Afghanistan, Argentina, Cameroon, the Haut-Katanga district of the Democratic Republic of Congo, Zambia, and Zimbabwe.

Evidence from HRITF's studies in Argentina, Cameroon, Zimbabwe, and Zambia show that RBF interventions can be highly effective for increasing coverage of maternal and neonatal health services, and also improving the quality of these services. Some of these studies have also been found to be cost effective.[33] The findings of these studies show that RBF may support health system strengthening in general through encouraging more active supervision and monitoring, measurable improvements in community involvement, and increases in health worker and patient satisfaction. The impact on the broader healthcare system suggests that RBF interventions may in fact support broader population health beyond the duration of the study periods.

However, the indicators which incentives target should be carefully considered so as not to create perverse incentives or unintended consequences. For instance, in a P4P study in the Haut-Katanga district of the Democratic Republic of Congo, facilities were allocated to treatment groups which incentivized increased quantity of care and control groups which incentivised increased quality of care. The treatment group reduced the cost of healthcare services to attract more patients but did not actually attract more patients which resulted in lower revenues for these facilities. This then led to treatment facilities having 42% fewer resources and 34% less income for health workers compared to the comparison group.[33]

Health facilities tend to be the primary targets of interventions. Although incentives may eventually be paid to health workers, the intervention design usually involves payments to facilities or regional/provincial management as opposed to directly targeted at individual health workers. The basis of payment for these allocations and the level of autonomy which facilities have in determining how they are spent also vary based on the intervention design.

Despite their promise, there have been very few Impact Bonds (SIBs or DIBs) in the health sector in LMIC settings (only five have been launched in South Asia and Africa[19] of which only two have completed). Some of the early lessons that may be relevant from the four African-based SIBs in flight include[34]:

- Using the Government as an active stakeholder, given their gatekeeper role in public health systems, is vital to a public sector programme's success;

- Regular (but simple and efficient) monitoring and feedback on performance is a critical enabler to ensuring the project is able to adapt and respond quickly enough where issues are found;
- Involved staff need to be trained and supported to understand and use the data in their day to day management of the project;
- Non-remunerated/financially-incentivised metrics tend to fall down the priority list. Therefore, the design of the incentives is important and important support functions should not be disincentivised;
- Alignment of the outcomes funders and service providers on the importance of the project is valuable.

However, the use of Impact Bonds (SIBs / DIBs) in public health systems is still relatively new and as these projects complete there are likely to be many more lessons to share.

With this context in mind, we delved into an extensive set of individual healthcare case studies, to truly understand what explains whether and how RBF does and does not work in achieving what it sets out to achieve (with consideration both between and within studies' results), and hence what guidance may be garnered for education.

3.3 How are education and health systems similar?

The education and health sectors are often compared to one another because they both form part of the "social services" sector.[14] They are both considered public goods, in that they not only confer benefits to those who receive health or education services, but society in general benefits from having a healthier and more well-educated population. Basic education and healthcare are also considered a human right. Due to the public good and rights-based entitlement to these services, both health and education are often provided by the State, at least at some basic level such as primary healthcare services and basic education. The provision of these services also becomes more expensive as they become more advanced, for example hospitalisation or university education.[14] As a result, some Governments choose to limit their financial support for these higher level services, limiting access for those who cannot afford private insurance, access personal loans or pay out of pocket.

Socioeconomic factors are major determinants of both health and education outcomes and therefore limiting access based on ability to pay has significant impacts on population health and economic growth. Although inputs such as well trained teachers and clinicians, and adequate facilities and equipment, are important for producing good health and education outcomes, social determinants, such as income, access to adequate nutrition, and physical and social environment, strongly influence education and health outcomes.[35–38] As a result, addressing factors outside the health or education system is necessary for addressing issues of equity in outcomes, making it difficult to create neat interventions that only target their respective sector.

Where tax revenue is used to fund a public health or education system, tax payers cross-subsidise non-tax-payers to ensure a reliable and quality public service. However, high income groups tend to access private health or education, due to perceptions of lower quality in the public sector. Therefore, these groups often have better health and education outcomes as a result of both better social determinants of health/education and the ability to pay for additional (or supplemental or preferred) services in the education and health sectors.

Both the health and education sectors are often characterised by strong unions or professional associations which protect the interests of healthcare professionals and educators, but can often be hostile towards the use of incentives to control the cost or quality of how these services are provided, and any new accountability mechanisms to monitor the impact of interventions.[14]

3.4 Identifying common incentive problems in health and education that RBF might address

The gap between best practice and available supply: While training may be of high quality, often public health or education settings are rife with inadequate infrastructure and supplies. This can create quality issues and lead to a decrease in staff motivation due to the frustration of being unable to provide adequate services with the available tools. This relates to structural quality.[39]

The salaried model can disincentivise efficiency and effectiveness: Salaries tend to be fixed and predictable in government jobs, making it difficult to incentivise efficient behaviours (salary does not change if outcomes improve or decline).[40] Traditional models of performance-related bonuses have fallen away in many public sectors due to financial constraints.[41] This has not been replaced by non-financial performance incentives.[42]

Principal-agent problem: Agency theory is usually used to explain how RBF is meant to effect change.[43] In this case the principal is the payer (e.g.: the Ministry of Health or a donor organisation) and the agent is the health facility or health worker. Agency theory assumes that the interests of the payer and the agent are not aligned. It is assumed that since most health workers in LMICs are salaried, public sector employees, and management structures within the health system are often centralised and weak, health workers have little motivation to work hard and provide quality care.[43]

In education, principal-agent relationships are complex, involving a number of layers: citizens (voters, parents, students), the state (executive, legislative, fiduciary), administration (national, provincial, district) and frontline providers (schools, teachers)[44]. Because of the potential for incoherence to arise in any one or more of these sets of principal-agent relationships, even the most well-designed intervention programs may fail.

The multitask problem: Health workers may focus less on tasks that are not being rewarded.[39,45] Where outcomes are multidimensional (for example, improved quality of care in maternal and neonatal health), incentivizing certain dimensions (for example, taking time to explain the importance of breastfeeding or prescribing iron supplements) may result in health workers focusing on those tasks which are incentivized,

but neglecting those which are not incentivized which may result in poorer health outcomes overall, despite seeing improvements in targeted indicators.[45] Education is prone to the same problem, which is one of the reasons why interventions are increasing focused very specifically on improving learning outcomes, as discussed above[20].

The know-do gap: This represents the difference (or gap) between the knowledge of what health workers should do versus what they actually do for clients.[46] This relates to process quality.[39]

3.5 How do education and health systems differ?

Although there are many parallels between the education and health systems, it is important to note the clear differences which impact how RBF interventions need to be designed. These differences influence the type of interventions required to have an impact on outcomes. The differences between the sectors should also be considered when comparing the relative effect sizes of the outcomes, as outcome measures in these different sectors may not be equally responsive or sensitive to interventions. We describe some of the main differences under the following broad themes:

Education is a more predictable service to run than health

- The number of learners each year is predictable and the education sector can plan with accuracy the inputs and resources required to run an education service. Health is a far more unpredictable environment, particularly at the level of hospital care, as one cannot predict who will need healthcare in the year with certainty.[47] Although the COVID-19 pandemic is a most exceptional event, it illustrates the point that specific demand for health services is unpredictable.
- Furthermore, while people are likely to re-enter the health system several times over the course of their life, the education system is discrete and has a defined start and end point.
- Therefore, knowing how many teachers, textbooks and schools a system needs is both easier to determine and easier to forecast based on the population growth estimates available in all countries. This may make it simpler to quantify the impact of RBF incentives in the education sector, which is something the health sector has struggled with.

The human resource requirements in health are more complex and require a longer planning horizon compared to education

- There is more complexity in health's human resource needs than in education. Human resources for health (HRH) is its own field and requires sophisticated modelling and planning to ensure the resources match the need, given the long training pipeline for some HRH cadres. This same complexity is not true for education and the training path to becoming a teacher is shorter than many of the HRH cadres.[48] However, the skills required by lower-level workers in health are considerably simpler than those required by their counterparts in education. Thus, the structured

play-based activities undertaken by kindergarten teachers to develop the psychomotor, perceptual, cognitive and emotional capacities in young children, and the variety of exercises required to develop reading skills in the early grades require both a sophisticated understanding of educational principles and their systematic application.

- Because health systems have a mix of HRH, they can task-shift some activities to lower level cadres to produce efficiencies and free up higher level HRH for more complex work. However, there is a limit to the extent of task shifting that can take place and the mix of skills required to manage a health system effectively is more complicated, sometimes requiring incentives to keep HRH in the public service or creating contracting mechanisms to allow private providers to deliver public care.
- As a consequence of these two factors, RBF programmes that aim to focus on human resource related challenges in education will need to be substantially different from the ones tested in the health system.

There are more inputs and equipment involved in running a quality health service than there are in running a quality education service.

- Health inputs, in terms of supplies, such as medications, vaccines, the appropriate medical equipment (e.g. ventilators), and facilities are all needed to effectively run a health service. For the education sector, the main inputs are textbooks and stationery, which cost less and are less complicated to secure and ensure than the health inputs. Laboratories, well equipped kitchens and IT equipment and connectivity only become important at the top end of high school. However, in education the most important inputs are expert teachers, which are not produced in sufficient quantities in many LMICs, to a large extent because of cost.

Adjusting for socio-economic and social determinants for education and health is different

- For both health and education, social determinants influence wellness, ability to learn and vulnerability to disease, learning difficulties etc. Therefore, for both sectors social determinants are a key variable to keep in mind when designing programmes.
- However, in health, there has been a concerted effort to introduce community outreach cadres and activities that provide support to people in preventing ill health and promoting wellness. These prevention and promotion activities are generally low cost, and some with very high reward in terms of outcomes. For example, teaching mothers how to make oral rehydration solution (ORS) for babies with diarrhoea using basic ingredients from the home has been a low cost, high impact intervention to account for the social determinants of health. 90% of under 5 deaths from diarrhoea could be avoided if ORS was used.[49]
- In education, the social determinants are not easily overcome through the same kinds of community interventions. One of the central social determinants in education is the social capital

held by families. For example, an important factor in preparing their children for school, and supporting them throughout the process is the way in which parents engage with their children from the day they are born, with better-educated, middle-class parents reading stories to them, teaching them to recognise letters and words, feeding them information, taking them on educational outings and engaging them in thought-provoking discussion, activities which working-class parents are far less able to undertake.[50,51] The very strong association between socio-economic status and parental education, on one hand, and the learning outcomes of their children, on the other, is one of the earliest and most consistent findings in educational research, leading to the famous aphorism by Basil Bernstein that 'Education cannot compensate for Society'.

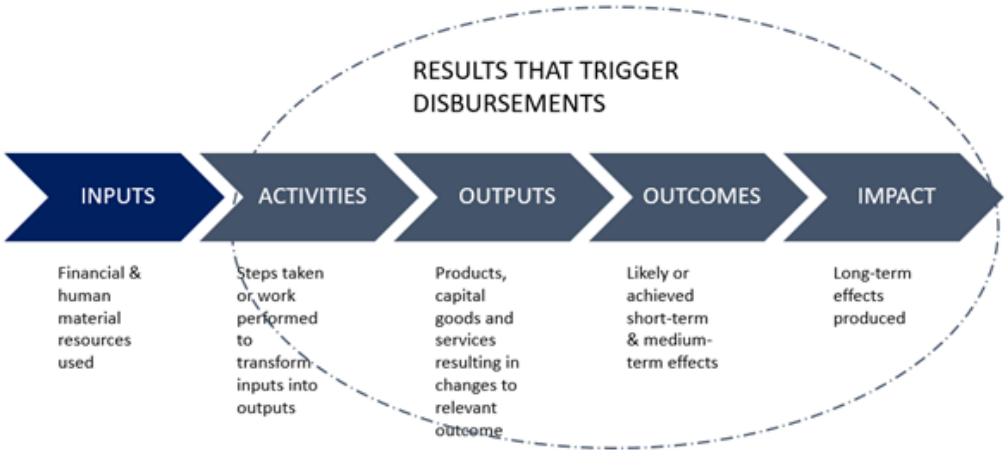
Having now provided some context as to what has already been done with RBF in health and education, and similarities and differences between the two sectors, as well as common incentive problems, we now turn to our approach to the detailed literature review.

4. Approach to Identifying research

4.1 Identifying relevant research on RBF in health and education

The literature review includes both peer-reviewed, scientific publications (qualitative, quantitative and/or mixed methods) as well as grey literature where it comes from a reputable source. We used a conceptual framework to guide our data collection and analysis process. The framework was derived from the OECD who developed this framework after reviewing the RBF literature (see Figure 1).[52] This framework can be applied at the different levels at which payments can potentially be triggered: national, local/district, and facility/school-level.

Figure 1: conceptual framework[52]



We have slightly adapted the OECD framework to include 'impact' in the circle of factors that can trigger disbursements, this was to allow for inclusion of social and development impact bond literature in our review. We considered the following terms as within the RBF body of knowledge and worthy of review:

- Payment by results;
- Payment for results;
- Results-based lending;
- Results based finance;
- Results based aid;
- Performance driven loans;
- Performance based aid;
- Cash on delivery;
- Output-based aid;
- Social impact bonds;
- Development impact bonds;

- Pay for performance; and
- Pay for success.

We used accredited search engines such as PUBMED and Google Scholar to identify relevant research. The PICO (Population, Intervention, Comparison and Outcomes) framework (a commonly used framework in health related research) was used to guide the search.[53] We used the following search terms, but where the term RBF is used we also cross-checked against the alternative terms listed earlier in this section:

Population

- Healthcare workers/teachers
- Health/education facilities
- Hospitals/high schools
- Clinics/primary schools
- Health units/ECD
- Health/education sector
- Health/education department
- Ministry of health/education

Intervention

- Payment by results;
- Payment for results;
- Results-based lending;
- Results based finance;
- Results based aid;
- Performance driven loans;
- Performance based aid;
- Cash on delivery;
- Output-based aid;
- Social impact bonds;
- Development impact bonds;
- Outcomes based financing;
- Pay for performance; and
- Pay for success.

Comparison

- Fee for service;
- Capitation OR per capita;
- Salary;
- In-put budget;
- Line-item budget;

- Global budget;
- Per diem;
- Case-based OR diagnosis-related groups

Outcome:

- Quality of care/teaching
- Quality of health/education services
- Quality of service provision in healthcare/education
- Health worker/teacher training
- Health worker/teacher knowledge
- Outcomes
- Public sector
- Structural quality
- Process quality

The search combined the terms in the PICO framework using Boolean terms.

4.2 RBF interventions targeting structural versus process quality

In healthcare clinical quality is usually measured by assessing the level by which health improvements have changed due to the service provided, but this is often too challenging to do.[39] Therefore, structural and process quality are measured instead.[39,54] Structural quality relates to the quantity and quality of health inputs such as drugs, equipment, and available staff.[39,54] Process quality relates to the way in which health services are provided, such as health procedures, diagnostic screening, and education provided to patients among others.[39,54]

Table 1: Examples of structural vs process quality indicators in healthcare and education

	Healthcare indicators	Education indicators
Structural quality	Doctors, nurses, CHW, other Human resources for health Hospitals, PHCs Medical equipment for health facilities Drugs	School managers (principals), Teachers, teaching assistants, ECD providers Schools Teaching equipment for schools (eg: teaching guides) Textbooks, student stationery
Process quality	Healthcare provided optimally and safely following the standards of service delivery through technical and non-technical performance	Teaching provided optimally and in a safe environment where the curriculum is taught and adequately covered by qualified and competent teachers. This can be assessed through standardised/national test scores and measuring student ability in reading, comprehension, and numeracy.

Assessment of these types of studies and possible relevance for the basic education sector will be furthered through our detailed literature review.

4.3 What kind of results have been incentivized in the health sector and how have these differed from those in education?

Table 2 below lists examples of the types of results which are targeted in education compared to health based on the level at which the disbursement would be triggered, as described in the conceptual framework of Figure 1.

At the input level, appropriate staffing and the availability of supplies tend to be targeted in both sectors. In terms of activities; productivity, reporting and student/patient support are targeted albeit in different ways based on the requirements for each sector. In healthcare, taking specific measures is an important activity for tracking health status, along with other routinely collected measures such as vaccine coverage, and indicators of patient satisfaction. Patient satisfaction measures are important to collect in healthcare as they explain the willingness (or lack thereof) to engage with the health system which ultimately impacts the outcomes of interest. At the health outcomes level, the results targeted tend to be related to usage of health services, mortality rates, and behaviour change (for example, using modern contraceptives and in-facility deliveries) which result in improved health outcomes.

Table 2: Illustrative examples of results in healthcare and education (see also Figure 1)

Level of incentive	**Illustrative results in healthcare**	**Illustrative results in education (with a focus on improving learning outcomes not enrolment)**
Inputs	• Adequate staffing (in terms of both quantity and skill level), • Availability of medical supplies and functioning equipment, • Reduction in stockouts of medicines and supplies at the facility.	• Fundamental inputs: adequate numbers of well-trained teachers, classrooms, books, WASH facilities, etc. • Intervention inputs in education - textbooks, for example - generally accompany other components, such as teacher training.
Activities	• Improved health worker productivity • Complete reporting • Staff trained to provide care to specific population groups based on specialised need • More mothers receiving counselling post birth • Proper case reviews after patient deaths conducted, • Increased number, and quality of prenatal care visits, • increased number of children with a complete	• Teacher education: Initial Teacher Education (ITE) and Continuing Professional Development (CPD) • Assessment of pupils and / or teachers • School feeding • Scholar transport • School fee offsets • Remedial education • Community-based education

Level of incentive	Illustrative results in healthcare	Illustrative results in education (with a focus on improving learning outcomes not enrolment)
	record of preventative health check-ups.	
Outputs	• Improved birth weight measures, • Increased vaccine coverage for mothers and babies, • Decreased waiting times at health facilities, • Increased patient satisfaction rates • Increase in use of modern family planning methods • Increases in the use and quality of prenatal care • Increase in use of antenatal care services • Increased number of deliveries that take place in health facilities	• Teacher and pupil attendance • Time on task • Improved teaching • Teaching at the right level
Outcomes	• Reduction in neonatal mortality, • Improved maternal, neonatal and child health outcomes	• Increased school enrolment • Improved equity (for example, with respect to the education of girls, rural communities and children from poor homes) • Improved learning scores • Reduced school drop-out rates / increased school completion rates

4.4 Summary of approach to evaluating studies and their relevance for education

After we had collated all relevant literature, we identified common themes and evidence under each of the processes in the results chain shown in Figure 1. We approached the question of whether the findings might be useful for the education sector in a broad fashion at first, learning what we could from the identified papers. After reading the papers, we found the following thematic areas (non-exhaustive list) to be most pertinent for the education sector (Figure 2). We discuss these in the next section.

5. The main and most relevant findings from health RBF literature for education

Our literature review confirmed that there is no one-size-fits-all solution for the way in which incentives should be designed to ensure that a health RBF program is effective. Incentives must align with performance objectives, but also consider a range of contextual factors that may impact the effectiveness of the intervention.[45] In addition, in order to ensure the sustainability of any RBF program which is proposed, incentives should be:

- **Fiscally prudent**: they should be sustainable beyond the life cycle of the pilot if it is found that the RBF mechanism is effective;
- **Simple to administer**: their implementation should not be onerous and should be aligned with national policies and regulation; and
- **Improve equity in access to quality healthcare**: they should strengthen, rather than distort access to quality health services.

In the section below, we describe the most relevant findings from RBF in health, that are relevant for education, with sight of these initial overarching points from the literature. We then consider how each of these lessons from health may be applicable in the education sector, with particular attention to improving the quality of primary schooling.

5.1 There must be clearly communicated coherence between RBF interventions and local policy contexts and objectives

Local context must be considered when designing and implementing RBF programs. It is important to acknowledge that although similar terminology is often used to describe types of interventions, such as "results-based financing", "pay for performance", or "performance-based financing" among others, there is no standardisation or uniformity in how (or where) these interventions are designed or implemented. As Kovacs et al (2020) describe, there is no "blue-print" for how RBF programs, specifically P4P schemes, are designed to be used.[55] This variation is often explained by different policy objectives, "conditions on

the ground" when the intervention is implemented, and organisational factors which influence how these interventions are designed. Even where similar interventions are implemented in similar settings, implementation fidelity may vary.

As discussed in section 3.1, RBF modalities have been applied in education by means of incentives to teachers, schools and families, but have generally not been utilised in pedagogically focused programs, such as Structured Pedagogy (SP) or Teaching at the Right Level (TaRL). Although such designs have been extensively applied using traditional funding mechanisms and found to consistently raise learning outcomes (see Appendix 1), the large majority of such interventions studied to date have been small in scale and not subject to any test of external validity. This is an issue that would be valuable to consider for future RBF studies on improving teaching and learning, because without external validity the evidence available to policymakers is limited. **Workers must be aware of how their performance is linked to local policy agendas and goals.** In a study of local stakeholder perceptions of PBF, Paul et al (2014) compared two RBF pilots in Benin, one implemented by the World Bank and the other by the Belgian Development Agency.[56] They found that the staff at health facilities did not understand how the RBF intervention linked with ongoing processes and reforms, particularly how it integrated with the universal health coverage agenda. As a result, RBF was not "owned" by health workers as a means through which their performance could be improved, but rather seen as a donor priority which added to their workload without much benefit.[56] The implication of this finding, equally applicable in health and education, is that care must be taken when introducing interventions to ensure that workers are aware of the extent to which the reforms conform to policy and are therefore intended to assist them to do their jobs. Therefore, RBF programs must consider the policy contexts in which these projects operate and ensure that health workers are informed on the coherence between the outcomes which their projects aim to achieve and national policy objectives. This as a strong theme in education as well, as discussed below.

The Chilean education reform example is particularly noteworthy in this regard (see appendix 1), in that the educational reform program was part of a government-wide initiative aimed at stimulating a package of economic growth, educational outcomes and social reform. According to the 2018 World Development Report[57] successive negotiations between the government and the teachers' union built broad support for a series of reforms that adjusted the working conditions of teachers to improve their overall welfare, while linking pay and career development more closely to performance. A number of authors warn that RBF initiatives can only be successful if buy-in is obtained from both government and teachers, and the case of Chile makes the point most cogently. The negotiation left both parties with gains: improved conditions for teachers linked to accountability for performance.

Public financial management (PFM) processes influence the success of RBF schemes. It is important to note that RBF schemes operate within the context of the particular health system. Specifically, the financing mechanisms and PFM processes will directly influence the success and sustainability of any RBF or P4P schemes.[58] RBF programmes are designed to influence the supply side of the health system to behave in ways that improve quality and health outcomes. In order to truly incentivise this, the supply side must have sufficient autonomy both in ***when*** facilities/providers are paid and ***what*** funding is spent on.

RBF may be more challenging to implement when decision making and financial management are centralised. While many LMICs have experimented with financial management decentralisation, it often

does not extend to the cost centre level (i.e. the health facility) which then means the RBF scheme is often reliant on district-level management teams/offices.[59] Furthermore, even where there is financial decentralisation, it is often not given with the requisite decision-making powers, meaning that health facilities are not able to direct their own activities sufficiently.[59,60] Therefore, RBF schemes that are targeted for regions with centralised financial management or decision-making are less likely to be able to actually incentivise the supply side of the health or education system, as they do not have the autonomy or flexibility to do so.[61] As a result, RBF schemes in this context may be less effective.

In a study of the role of the health system context in the design and implementation of performance-based financing (PBF) in the Ivory Coast, the authors found that the limited improvements in quality and outcome measures could partly be attributed to "constrained decision making space" available to health providers due to the country's centralised health system structure.[61] Health facilities did not have the autonomy to hire or fire health workers or to allocate their own non-PBF budgets, and these political economy contextual factors, which are not directly influenced by the PBF scheme, influence health quality and outcome measures.

Research on the political economy of education in developing countries is scarce, and even less research on how to implement interventions in challenging political contexts.[62] However, there are recent studies to suggest that varying political commitment in countries such as Bangladesh, Cambodia, Ghana, Rwanda, South Africa, and Uganda, negatively affects reforms aimed at improving the quality of education.[63] The authors found that despite impressive expansion of access to primary schooling, learning outcomes remain poor in these countries. For example, Pratham,an Indian-based NGO which aims to improve the quality of education in India, has shown impressive results. Despite these successes, replication across the country and the sustainability of such initiatives remain a challenge[64], in part due to a lack of political will and commitment. According to Bano and Oberoi (2020), NGOs such as Pratham are successful due to their networks of volunteers that enable such organisations to tailor their implementation according changing political contexts, as opposed to ceasing operations when state actors withdraw their support.[64]

Reflections by Banerjee and colleagues on the efforts of the TaRL initiative to move from proof of concept to widespread adoption across a number of Indian states reveal several hurdles to the implementation of donor-initiated interventions in education[22]. Chief among these, in the words of Banerjee et al (2017:95) is to find answers to the question: "How do you get a bureaucracy to make a common-sense change that has a very strong chance of being beneficial", where the chances of success have been established by means of rigorous research[22]. **Discrepancies between implementation sites should be accounted for when setting performance targets.** Perceived differences in premiums granted between World Bank and Belgian Development Agency facilities also created a feeling of unfairness among staff at facilities which received lower premiums.[56] In addition to this, discrepancies between facilities were not accounted for, meaning that facilities which initially did not meet national standards due to lacking the required resources or equipment were assessed by the same criteria as those that were better equipped.[56] This meant that regardless of their best efforts, these poorer resourced facilities were penalised, and this resulted in a perception among health workers that RBF was unfair.[56] This also shows that there needs to be improved coordination among donors where similar programs exist so that design differences between these projects do not engender negative attitudes among participants. Lastly, if factors which influence the ability of facilities to attain targets differ across facilities, such as resourcing, then equity weights should be applied so that indicators and incentives should be adapted to account for this.

Multiple studies reviewed included equity weightings or bonuses in the RBF calculation for rural areas, such that facilities in rural areas received higher incentive payments relative to urban areas. This was, for example, the case in the Mozambican program mentioned above, as well as in a RBF program in Cameroon[65] which also focused on mother and child health. The latter intervention (in Cameroon) aimed to improve service quality, measured through the use of a checklist and patient satisfaction surveys.[65]

Context should also be accounted for when translating learnings from the health sector to education. These contextual factors are important to note when considering what can be learned from different studies within one discipline, such as healthcare, but even more so when considering how relevant these learnings are to other disciplines such as education. The consideration of the local context in which these reforms take place is paramount, irrespective of sector. The two interventions in education which have demonstrated significant learning gains in early grade reading and gone on to adoption at scale by government have obviously taken meticulous account of context, both politically and educationally. See appendix 1 for a discussion of the Tusome and TaRL projects in Kenya and India, respectively.

The health sector has more experience with alternative reimbursement mechanisms (ARMs) compared to the education sector. In Health, incentivising quality and performance through ARMs is no longer novel.[66] In public and private health systems in many high-income countries and more recently, even in LMICs, incentive-based payments such as diagnostic related groupers (DRGs), global fees and results-based financing schemes are used- all with the intention of making healthcare spending as efficient and effective as possible in order to improve health outcomes.[67] This longer-term experience means that health sectors are likely more adept at managing ARMs than the education sector would be. However, having multiple ARMs running within one health system can be complex, and managing this complexity requires a fairly mature system with strong leadership and governance capabilities.[68] Where it is poorly managed, the ARMs could work sub-optimally or actively work against one another if the incentives are not well aligned.

Alterative reimbursement measures have not been much studied in education, where funding tends to be disbursed centrally. Where funding is decentralised it may be done by outsourcing provision to private sector providers, such as happens, for example, in Kenya and other African countries by means of government subsidies to low-cost private schools [69].

High proportion of vertical donor funding can lead to fragmentations, lack of funding coordination, and make programs unstainable. The proportion of donor funding in health sectors in LMICs is generally high, which can cause some fragmentation given the different financing flows.[70] Understandably, donors are usually funding to improve a particular outcome and therefore they want to ensure their funding is targeted for maximum impact. However, if there is no process for a ministry of health to appraise, coordinate and direct donor funding in a country, these schemes can land up causing confusion and working against one another.

RBF programmes in LMICs are almost always donor initiated and/or supported and domestic funds are rarely devoted to these interventions.[15] This raises concerns around the ownership of these programmes and their sustainability in the long term.[15] Therefore, a key lesson for the education sector lies in the setting up of a robust appraisal process for pilots and trials of RBF schemes, and ensuring their

sustainability by establishing local ownership of projects and domestic funding of projects, at least in the long term.

5.2 Understand the difference between indicators which are influenced by process vs structural quality

Quality of care or service in healthcare can be divided into factors which influence structural quality and those that influence process quality.[39] Structural quality indicators relate to *inputs* which influence the quality of service, whereas process quality indicators are those that relate to *how* quality healthcare is delivered. It is important to understand whether the indicators which are targeted for improvement in RBF interventions are influenced by process or structural quality so that incentives can be targeted accordingly.

Results are better when the provider has direct control over what is being incentivised. In their review of the design of P4P schemes in LMICs, Kovac et al (2020) found that healthcare visits (83%) and process quality (66%) were the most commonly targeted measures for improvement. Some structural quality measures such as ensuring the availability of waste management and infection control, and management practices including the upkeep of medical records, were also targeted in 27% of studies, but health outcomes and efficiency were targeted least.[55] Kovacs et al (2020) argue that where the health provider has control over what is being incentivised, results tend to be better. So, for instance, since health providers have direct control over process quality, this is what should be incentivised, and although process quality may influence health outcomes, the provider has less direct control over health outcomes.[55,71] The same is true for structural quality, although this is more under the influence of health facility managers as opposed to individual health workers.

For example, a health worker may be incentivised to reduce cardiovascular mortality rates, but despite conducting the appropriate screening tests and prescribing the correct medication (process quality indicators), their patients may refuse to make the lifestyle modifications (such as diet and exercise) to reduce cardiovascular mortality rates.[71] Therefore despite their being no problems with process quality, outcomes may remain unchanged. **Clearly identify the problem and target that specifically.** It is also important to have a clear understanding of what the underlying problem is which is being addressed, so that the incentive can be targeted directly at the mechanism through which the financial incentive can help. Therefore, if the problem is that there are constant drug stock-outs, the financial incentive should target the supply chain mechanisms which result in this as opposed to, for instance, health worker motivation.[72]

In education, as noted in the REACH report[73] for instance, a basic problem facing many education systems is a lack of properly trained and effective teachers. Seen from this perspective, a likely explanation for the mixed but generally disappointing results of the effects of teacher incentive schemes on learning outcomes might be that incentives may influence teacher effort (more of the same) but may not result in improved pedagogy (qualitatively different). Thus, where poor teacher effort is a major problem, incentives may lead to improved effort and hence better outcomes, but where teacher capacity or ability

(process quality) is the major cause of poor outcomes, then incentives are less likely to improve the situation: if teachers don't know how to teach better, then offering them money won't help. Therefore, RBF schemes must be designed to target improved capacity by providing opportunities for teachers to act on incentives through, for example, training opportunities or use of teacher guides, which can often improve the effects of incentive schemes on learning. This is one example of how clearly identification of the cause of a problem is important in determining intervention design.

Quality of care is complex to measure. Quality of care is often estimated using composite measures or indices which involve the weighting of various different components.[74] For instance in a study looking at the impact of P4P on maternal and child health services in Rwanda, the quality score used to determine the amount of money a facility would receive as part of the P4P scheme was an index with both process and structural quality components. Table 3 below shows how each of these components was also broken down into measures such as administration, the curative care provided, immunisation, family planning, and tuberculosis services, among others. Facilities were scored on each of these measures, and the weighted sum of these measures is the score allocated to each facility. In this study, P4P was found to improve the quality score, therefore showing evidence in favour of P4P improving quality of service in this study.

Table 3: Services and weights used to construct quality score for P4P formula in Rwanda

	Weight of service in quality index	Share of weight allocated to structural measures	Share of weight allocated to process measures	Means of assessment
General administration	0.052	1.00	0.00	Direct observation
Cleanliness	0.028	1.00	0.00	Direct observation
Curative care	0.170	0.23	0.77	Medical record review
Delivery	0.130	0.40	0.60	Medical record review
Prenatal care	0.126	0.12	0.88	Direct observation
Family planning	0.114	0.22	0.78	Medical record review
Immunization	0.070	0.40	0.60	Direct observation
Growth monitoring	0.052	0.15	0.85	Direct observation
HIV services	0.090	1.00	0.00	Direct observation

	Weight of service in quality index	Share of weight allocated to structural measures	Share of weight allocated to process measures	Means of assessment
Tuberculosis services	0.028	0.28	0.72	Direct observation
Laboratory services	0.030	1.00	0.00	Direct observation
Pharmacy management	0.060	1.00	0.00	Direct observation
Financial management	0.050	1.00	0.00	Direct observation
Total	1.000			

Although composite measures capture complexity of quality measurement, they may mask improvements in individual indicators. However due to the complexity of these measures, the impact of RBF on quality is not always observable through these indices. A study of the impact of RBF on maternal and neonatal health in Malawi showed that although improvements in individual quality indicators/measures were detected, there were no significant improvements in the composite measure of care quality.[75] Quality of care is multidimensional which makes it a complex measure to capture using coverage indicators alone,[13] therefore a mixed methods approach (using both quantitative and qualitative research methods) to assessing quality is more beneficial than simple checklists as they provide a more comprehensive understanding of the quality adjustment process.

Due to the complexity of these measures, it is important for health workers to understand how the inputs measured feed into how they are being evaluated, in order to ensure the success of the intervention.[13] Therefore it may be advisable to first begin with structural quality indicators that relate to inputs and then progress to more complex process quality measures of clinical care to develop a more thorough understanding of how RBF is ultimately meant to improve quality of care.

In the education sector the predominant focus of intervention programmes is aimed at improving teaching quality, which is a process quality measure. This follows the realisation of a growing 'learning gap' in LMICs, characterised by high levels of primary school participation and poor learning outcomes in the foundation disciplines of literacy and numeracy. Two design types stand out as being successful in raising test scores in this context: Structured Pedagogy (SP) and Teaching at the Right Level (TaRL). Neither approach has, to date, made use of RBF modalities. Both designs are relatively complex, depending on one or other combination of a number of structural and process quality components: the provision of materials, teacher training, monitoring and support and classroom reorganisation. However, both the target of such programs (teacher pedagogy) and the measure of success (test scores) are simple and explicit.

5.3 Incentivise agents involved at every organisational level relevant for achieving the desired results

Where indicators of interest involve procurement along a value chain, all agents along the value chain should be incentivised. In a study which used P4P incentives to improve the availability of medication and reduce stock outs of essential medicines in Tanzania, it was found that incentivising the district *and* facility management meant that everyone relevant to the procurement of drugs was working toward the same goal.[72] Verification of stock outs happened at the district level, therefore district health managers were involved in avoiding stock outs at facilities. The authors also found that incentivising availability and avoidance of these medicine stock-outs may be effective alongside broader incentives for service provision, especially if these have complementary aims, but it may also be useful as a stand-alone incentive program.[72]

Incentives should be targeted at individual agents where individual effort is important. Since most of the health literature focuses on RBF at facility level (which is then filtered down into payments to staff members) there are fewer health studies on demotivating factors at individual staff member levels. One such study which was focused on healthcare community worker cooperatives did however highlight this, stating the fact that each community health worker's effort was largely unobservable whilst payments were made at the cooperative level (with these then being filtered down to the workers), meant that free-riding may have been incentivised.[76] This program did not produce results, for a myriad of possible reasons mentioned by the study authors, including this feature.

Conversely, an RBF project in government primary schools in Andhra Pradesh utilised two types of teacher performance incentives; a group bonus for improved outcomes at the school level, and an individual teacher bonus for improved outcomes by their students.[56] Muralidharan and Sundararaman (2009) found positive effects of both types of bonus payments on maths and language outcomes.[77] While the effects were similar in the first year, in the second year the schools with individual-level incentives performed better. In parallel, the study also randomly assigned two other sets of schools to receive increased inputs, and found that the schools with incentives outperformed those with just increased inputs. At the end of two years of the program, students in incentivised schools performed significantly better than those in control schools by 0.28 and 0.16 standard deviations in math and language tests respectively; individual incentives resulted in higher gains in student learning than group-based incentives.[78]

Incentives which reinforce norms of professional behaviour can enhance intrinsic motivation. Extrinsic motivators, which are central to PBF models, such as financial incentives, increased supervision, external verification, and feedback processes which are often built into PBF performance contracts, help to align interests of health funders and health providers, and thus overcome the principal-agent problem (described in section 3.4).[43] However, in the health sector in LMICs, where working conditions are often poor and salaries are low, altruistic motives or "intrinsic motivation" are also important influences for health and educations workers' behaviour and attitudes towards their work.[43]

The Andhra Pradesh program drew for its design from the psychological literature on incentives, which suggests that extrinsic incentives that are perceived by workers as a means of exercising control over them

are more likely to crowd out intrinsic motivation, while those that are seen as reinforcing norms of professional behaviour can enhance intrinsic motivation. Thus, the way an incentive program is framed can influence its effectiveness. The program was careful to frame the incentives in terms of recognition of excellence in teaching as opposed to framing the program in terms of "school and teacher accountability.[77] This links strongly with findings from the health literature on the importance of considering who is incentivised, as well as how it is incentivised.

Separation of functions is critical to target incentives correctly and monitor performance accurately. The value of a purchaser provider split (PPS) within health systems is widely accepted and a common goal as countries pursue universal health coverage.[79] The purpose of a PPS is to centralise the pooling of healthcare funds, the design of the benefit package and the basis of payment to healthcare providers and in that way, bring down the cost of healthcare, and improve the quality of care, through more strategic purchasing and economies of scale.[79,80] A PPS is able to improve accountability of providers through the way they reimburse, and by separating out purchasing from provision, a purchaser is able to build in payment incentives, such as RBF much more simply.[81] Independent verification of results and monitoring of performance should also be separated out from those who stand to gain from the incentive payments (or the non-payment of incentives). Valuable RBF schemes therefore require intentional separation of functions in order to verify achievements.[82]

A Bolivian study looking at the impact of results-based management programmes of a health services network in the El Alto municipality provides an illustration of the value of the PPS.[83] As part of an experiment in the reform of the health system, the municipality entered into agreements with a non-governmental organization (NGO) to take over the overall management of one of the El Alto health services networks.[83] The contracting of management was based on achieving process and outcome indicators and was therefore considered as a PBF scheme. One year after implementation, the preliminary results of the experiment showed improvements in the utilisation and occupancy rates of secondary hospitals, increased outpatient consultations and institutional deliveries, and improvements in client satisfaction.[83] These improvements were attributed to the changes in organisational management which created the separation between the purchaser (the NGOs) and the providers of services, in combination with the results-based management and improved community participation.[83]

5.4 Incentive payment design should account for how incentive size, level of effort, and payment frequency impacts outcomes

Health studies highlight the importance of carefully designing RBF programs with suitable payment designs. Payment design may include consideration for the suitable size of payments (with consideration of how much additional effort is required by a health worker to complete an incentivised action) and the frequency of payments. It may also include consideration of how money may be used, whether there are differences in the level of incentives provided to different facilities or people (for example in more or less resourced areas), whether payments form part of salaries or bonuses, and whether incentives are in the form of additional payments or penalties.

Things that are easier to do are easier to incentivise. Regarding the size of payments (and consideration of marginal effort), an RBF program that focused on improving mother and child care at the facility level in Rwanda was designed to incentivise items such as prenatal care visits and the quality of prenatal care (using a quality index score).[84] A study of this program concluded that it had the greatest effect on those services that had the highest payment rates and that needed the least effort from the service provider.[84] On the contrary, a RBF program in Mozambique[23] found that indicators incentivised were not sensitive to price but rather inversely correlated with the amount of marginal effort required. In other terms, if something is easy to do, it may be more easily incentivised. This may especially be relevant where small changes in health or education service provision have proven to lead to large results, relative to the amount of effort required.

The size of incentive payments matter, up to a point. Notably, different studies have found that payments that are either too low or too high may be sub-optimal in achieving the aims of RBF. One study showed that when payments were 2.5% of providers' annual salary, performance effects were not seen, suggesting that this incentive was too low to change behaviour.[85] Other studies state that larger incentives lead to greater performance, but after an initial boost, performance declines again. This may be explained by the reference- or target- income hypothesis, which indicates that no further improvements will be observed once a certain level of income is reached.[86]

More frequent payouts lead to greater improvements in performance. Regarding frequency of payments, most health studies reviewed made RBF payments on a monthly or quarterly basis, generally with strict processes to ensure verification and payment on time. Individuals tend to discount future gains, therefore, providing smaller incremental payments could result in greater performance compared to once-off payments at the end of the year. This is an important learning for RBF design. Frequent pay-outs also increase incentive salience.[86]

Rewards work better than penalties. A common theme in all studies reviewed was that incentives were designed to be for performance, as opposed to penalties for lack of performance. This follows the well-established loss aversion phenomenon, where the degree of demotivation derived from losses may be significantly higher than the degree of motivation derived from rewards, possibly leading in RBF programs to unintended consequences. A related theme is that RBF programs in health focus on payments in the form of bonuses or "salary top-ups", as opposed to using RBF to form base salary payments. Due to the perceived difference of receiving an additional bonus compared to having the same value added to the existing salary, performance can be better incentivised if incentives are paid separately to regular salaries. However, this could create payment-related administrative complexity that some LMICs may struggle to cope with.[87]

Bonuses may be useful for incentivising quality and quantity targets. Health RBF programs that target both process quality and factors such as service volume or coverage simultaneously, also often incorporate both quantity and quality components in the calculation of payments. In the an RBF study in Cameroon, the payment calculation included an additional quality bonus that provided an increase of up to 30% of the total payment based on health service quantity (where quantity was focused on items such as number of antenatal care visits and facility-based birth deliveries, and quality based on a checklist that indicated if a health provider completed routine items such as basic health checks and necessary information gathering about the patient). A similar payment structure was in place in a RBF program in Burundi, where

a quality bonus of up to 25% was awarded to facilities based also on similar checklists (meaning that the total payment was calculated as a weighted sum of the number of provided services multiplied by their unit payment, multiplied by a quality bonus, where the latter had a range depending on a score obtained from a checklist assessment).[88]

Absolute targets foster healthy competition, while relative targets promote continuous improvement. Some programs have observed that if RBF is based on set targets, then once a target has been reached, further attempts to improve may not be made. Having specific targets based on the baseline performance of the provider may overcome this. A common feature in the health studies reviewed was that RBF payments were based on units of care or quality checklists in a continuous manner, as opposed to only being paid when a certain target is attained.[45] Similarly, absolute targets rather than relative targets that foster competition may be preferable and are generally used in health RBF. The certainty and transparency of absolute targets are considered more acceptable, as long as the goal remains achievable. Relative targets (in relation to other facilities) have the advantage of leading to continual improvements. However, this competition may have a negative effect on collaboration and could result in gaps in performance between different regions.[45]

Additional funding, improved oversight, and autonomy on how to allocate additional resources may be more important for quality improvement than performance targets. Although well designed performance linked incentives can lead to improved outcomes, a study of RBF in Cameroon shows that it is the additional financing and autonomy which is the determining factor in quality improvement as opposed to linking this financing to performance. In this study, RBF was compared to non-RBF controls groups (which received supervision and monitoring only) to isolate the impact of the performance linked financial incentive on quality of service and utilisation at health facilities.[65] The groups which received financial incentives both had large and statistically significant improvement in the availability of equipment and other structural quality measures compared to the non-financially incentivised control groups. However, there was no statistically significant difference in structural quality improvement between the group which received performance linked financial incentives (RBF) and the group which received addition financing which was not linked to performance. This suggests that providing more funding, along with supervision and autonomy of how these funds are spent, is necessary to improve structural quality even when payments are not linked with performance incentives. However, providing supervision and monitoring without increased funding, and the autonomy to allocate these funds, is not sufficient to improve service quality, as there was no statistically significant improvements in these indicators for the groups which received no additional financing but only supervision and monitoring.

In contrast to the extensive literature on incentive payment design in the health sector, there is a paucity of information on this issue in education. The REACH report describes mixed results regarding design details such as the size of incentives or whether they target individuals or groups of teachers, concluding that there is no conclusive evidence that proves that incentives work best at one particular level of the education system[89]. However, the report does emphasise the important point that perfecting incentives at just one level will be of little use if the incentives at the other levels are not aligned, requiring that designers have a clear theory of change, and be aware of the political economy of the environment into which an intervention is introduced. A second important conclusion of the REACH report is that, because

education initiatives require relatively long lead times to become embedded and to begin to affect outcomes, project timeframes of 4-5 years may be insufficient for the desired results to be achieved.

5.5 Increased feedback, supervision, and training are important for supporting performance improvements, but autonomy in spending may improve outcomes beyond incentivised targets

Increased supervision and financing are reinforcing with regard to performance improvement. RBF programs in health tend to include a supervision, monitoring, measurement, feedback, coaching, or support component. However, it is not always clear whether improvements in outcomes are attributable to this increased monitoring, support, and feedback, or rather the financial incentives, or both. A cluster randomised trial RBF study in the Philippines, which looked at the impact of performance bonuses paid to physicians on child health outcomes, found that children who were treated at facilities where physicians were part of the RBF intervention showed statistically significant improvements in general self-reported health measures and wasting over time compared to children in the control group.[90] However, incentives were only paid to physicians after 36-months, because health outcome data was only collected 2-years after the intervention.[90] Therefore the study authors hypothesise that measurement and feedback of physician care quality played a primary role in improving health outcomes, and the bonus payment acting as an "accelerator."[90]

On the other hand, an RBF program in Cameroon that formed part of our review included "systematic supervision of health facilities", defined as "regular supervision by an external supervisor from the district hospital team using a structured checklist and providing immediate feedback to facility staff on problems identified and potential solutions to improve service delivery."[65] This study highlighted the importance of this, but also concluded in this case, through the use of counterfactuals, that enhanced supervision and monitoring alone was not enough to improve service delivery, additional financing (whether traditionally or in the form of RBF) was crucial to this aim.[65] The authors describe these two elements, supervision *and* financing, as reinforcing. Little was mentioned in this study about the form of any further coaching. A similar finding emerges from the education sector: while there is little evidence regarding the effects of increased accountability on its own, it appears to be an important factor enhancing the success of programs which incorporate other measures, such as providing incentives to teachers to improve attendance[89].

Monitoring and support to teachers is a key element of the pedagogically-focused interventions (see Appendix 1: Examples of proven pedagogical approaches that have improved outcomes in education. Briefly, effective support to teachers is a package consisting of materials, on-site coaching and monitoring. During the experimental phases of such programs, external technical advisors provide coaching and monitoring services, and in those few programs that have gone to scale, government agents are involved in these services from the start. Given the complexity of raising teacher capacity and the lack of understanding by school principals of how to do so, those few interventions in education that have been successful at scale have been effected by incentivising an outside agent to initiate the process, provide intervention activities at teacher- and school-level, monitor progress and hand the program over to

government, as was the case in India[22] and Kenya[83], or been initiated within government itself, as was the case in Chile[57].

Autonomy on how incentive payments are spent may improve outcomes beyond incentivised targets. In addition to incentivising the correct actors, providing them with autonomy over how the incentives are spent, as opposed to this being decided by the payer, has been found in health programs to achieve better results as this improves intrinsic motivation of staff.[33] This was a theme observed in the learnings from the HRITF portfolio of impact evaluations as well. They observed that in RBF interventions where there was greater autonomy over how funds were spent, improvements were not only observed in the indicators which were incentivised, but also other health system challenges such as medication stock outs and non-incentivised conditions such as non-communicable diseases. In Argentina, an evaluation of the P4P program to incentivise increased enrolment in Plan Nacer found that the financial autonomy provided to facilities allowed for improved allocation of resources, which also had a positive impact on the health outcomes of patients.[92]

There are similar findings in the education sector, where the success of programs in which grants are paid to schools is associated with institutional autonomy. The Punjab province of Pakistan seems to have found an effective way to use incentives to improve learning outcomes, through the Pakistan Foundation Assisted Schools (FAS) program.[27] FAS is run by the government-supported Punjab Education Foundation (PEF), established as an autonomous non-profit body by the Punjab Education Foundation Act of 1991. PEF provides monthly subsidies to support low-cost private schools in the region, conditional upon them meeting specified target pass rates on standardised exams. The evaluation concludes that the key factors contributing to the program's sustainability and success include strict program requirements for participating schools and strict adherence to the RBF model, under which the threat to schools of losing the subsidy for failing to meet targets is real.

Increased autonomy initially requires increased support and regular feedback. One health RBF program in Benin highlighted the coaching and training component of the program, with authors of a study here stating that "It has been noted that facility staff and managers appreciate and are responsive to having detailed feedback on their performance, this is even more useful if accompanied by supervision (by district health authorities) and coaching (by the implementing agency) to identify issues limiting the facility's performance as well as strategies to address them."[93] The authors went on to state that the latter is a "key element of RBF because with increased autonomy, providers need more data and (at least initially) external support for decision making."[93] This study did not however quantify the effects of this, but rather emphasised that verification took up a great deal of time, leaving less time for these essential components.

In health there is less focus on coaching and more on bridging the know-do gap. Notable is that in many of the health studies reviewed there was often little to no mention of something akin to coaching and training alongside the financing, albeit with a higher focus on checklists of items for complete servicing and quality service provision, as was mentioned above in the programs in Cameroon[65] and Burundi[88]. This may be due to the fact that health staff are often sufficiently trained, and the main issue is translating their skills and training into actual practice (ie: the know-do gap). Although some RBF health studies do mention coaching and training as key elements, this is generally geared toward achieving the RBF results intended, rather than filling basic training gaps in workers' education.

5.6 Possible unintended consequences

Throughout all the above-mentioned design features of a RBF program, the health literature recommends that care is taken to avoid any unintended consequences. Unintended consequences may materialise in many forms, some of which have already been mentioned above. Of primary concern is that those incentivised may game or manipulate the system in order to increase compensation, and features of the design may have the counter effect that they produce inequitable results or are simply demotivating to those incentivised. We have already mentioned equity adjustments above.

An understanding of unintended consequences is crucial for knowing the true impact of an intervention. It is reported that although it is hypothesised that there may be numerous unintended consequences to RBF interventions, there is limited research on them.[94] However, it is important to lookout for them as they may have implications for quality of service outcomes, and the health system more broadly, outside of the outcomes which are being targeted. A thorough understanding of the intended and unintended consequences is crucial for truly knowing the impact of an intervention.[94] For instance, given that RBF in health does not always incentivise the full service package but rather a subset, provider effort may move from non-incentivized to incentivized services. This is observed in the example provided in section 5.4, where RBF was seen to have the greatest impact on the services which required the least effort.[84,94] Other unintended consequences include the oversupply of incentivised services at the expense of those that are not incentivised, manipulation of data, hiding stockouts of essential medicines, observing professional norms only during PBF supervision, undue penalisation (see discussion in section 5.1), work overload, and reduced job satisfaction among health workers.[15,94] There seems to be insufficient evidence regarding unintended consequences as a result of RBF programs in education to draw conclusions in this regard. However, the REACH report notes that this may be due to a lack of adequate monitoring mechanisms needed to catch such undesirable behaviour[89].

Audits and penalties are needed to ensure funds are appropriately used. Given the propensity for RBF to sometimes create perverse incentives unintentionally, it is important that verification is robust and the consequences for gaming or cheating are salient. Kovacs et al (2020) highlight the importance of having audits for RBF programs, with penalties in place.[55] For illustration of how to implement this, a RBF program in Mozambique that was focused on mother and child health and HIV care included penalties to dissuade inaccurate reporting.[23] In this program, health facilities reported on RBF indicators monthly and submitted aggregated reports, based on facility registers, on a quarterly basis. Data verification and payment cycles were then scheduled for every quarter and were conducted jointly by the program implementer and the provincial health department. In this process, verification was thorough; every indicator at every facility was verified, with standard operating procedures guiding this process. Reported outcomes were cross-checked with a source register book, after which facility register data were cross-checked with patient files and pharmacy records. A 10% variance between reported and verified reports for a particular indicator resulted in non-payment for that indicator. If a 10% discrepancy was found in more than a third of all indicators for a facility, then that facility forfeited all RBF payments for that quarter. These audits were carried out by multiple teams – the study team, the NGO implementation team (funded by the US CDC), and an independent auditor. Whilst it is not possible to know what the

counterfactual may have been in the absence of this robust system, all audits under this RBF program were reported as clean.

Verification can be done in a way which cultivates accountability and improved communication. Such verification processes may also be incorporated into a program in such a way that it fosters increased communication between those involved in achieving the aims incentivised. In a Tanzanian RBF program that aimed to incentivise the avoidance of medicine stock-outs, verifications at facility level were done by district managers, thereby having the added benefit of increasing district managers' involvement alongside the involvement of facility managers and frontline staff.[72]

Lack of contextual understanding and study durations that are too short may lead to irrelevant or invalid study results. A final and most critical unintended consequence may be that results are not relevant and valid; that metrics are myopic or misleading if context has not been fully understood at the start and during implementation of the program; and that results may slow or end if and when funding or the study efforts end. A 2017 Mozambican study highlights that it was one of the first to thoroughly assess impact changes of RBF over time.[23] Their conclusion was that, on average, RBF takes six quarters (18 months) before impact can be observed and sustainable.[23] Over this time period, health systems' actors can fully understand and internalise how RBF works, develop strategies to improve indicators, and organise systems and processes to achieve relevant and valid results.

The World Bank's Innovation Trust Fund Learning Portfolio has already highlighted other unintended consequences, such as dissatisfaction by staff due to reduced unit prices, and increased patient load leading to burnout. Other studies also discuss the problem of free-riding when incentives are paid to the group for outcomes that depend on individual effort. Many of the unintended consequences listed here can be avoided by through appropriate study design which foregrounds the findings described in the section 5.1 to 5.5.

5.7 A note on impact evaluation

Many of the studies reviewed in the health literature had extensive results, some set up detailed counterfactuals (as for example mentioned above in relation to the Cameroon study[65]), however understanding the full context is key to understanding results and knowing whether or not they are as a result of the intervention or other factors.

In the Mozambican study mentioned it was notable that some targeted care did not respond (for example HIV care, whilst MNCH care did).[23] The authors of a study around this program state that this was due to different starting points/ room for improvement, meaning sufficient funds for certain types of care to begin (in this instance Mozambique had concurrent programs for input and output based financing, and already had some HIV programs in place). In the absence of knowing this, understanding whether RBF does or does not work and why may be misleading.

A second example is the RBF program mentioned in Tanzania.[72] In this the stocking of certain medicines and equipment did respond to RBF, whilst others did not. The authors of a study around this program

attribute this difference to various factors including the extent to which certain medicines stocks were targeted but not necessarily linked to district managers' bonuses, the extent to which certain areas and facilities starting points were better and hence had less room for improvement in good stock practices, and the extent to which some items depended on donor funding, hence limiting how well RBF could work with these external factors. Some items, equipment in particular, also were very expensive and so stocking of these items lagged other items such as lower cost drugs. In these cases, incentivising the avoidance of stock-outs require consideration of all of these contextual factors. Exploring all of these factors was key to understanding whether this program worked or not, and why or why not, with detailed attention to different aspects of it (in this case, differences in stocks of different medicines).

Whatever the specific program design, a rigorous monitoring, verification, and evaluation system is essential for establishing impact and unravelling the mechanism of any change detected. An important example in the education sector in which this principle was not adequately addresses is afforded by DFID's Cash on Delivery Aid program in Ethiopia, GBP 10 million per year would be paid to the Ethiopian Ministry of Education, contingent on an increase in the numbers of students sitting and passing the Ethiopian General Secondary Education Certificate Examination (EGSECE) at the end of Grade 10 in the years 2012-14.[95] While the number of sitters and passers did increase in this time, an evaluation concluded that *"…none of the estimated impacts on the numbers of either boys or girls sitting the EGSECE were either statistically significant or reasonably attributable to the RBA pilot"* (ibid: iii).

The inability of the evaluation to attribute impact to the program is at least partly the result of an inadequate evaluation design which, due to cost considerations, precluded the use of a reliable counterfactual. This quote highlights an important lesson for interventions in health and education: "tracking the effects of any program requires a sophisticated verification system if all the consequences and the mechanisms for achieving them are to be fully understood."[55]

Rigorous experimental evaluation designs are key not only to establishing whether health or education interventions result in significant changes in the targeted outcomes, but also in attributing the causes of such changes. In addition, qualitative studies are important in order to understand the mechanisms underlying the changes.

6. Discussion

The aim of this report is to understand what can be learned from RBF as applied in the healthcare sector in order to inform RBF in the education sector, particularly in LMICs and with attention to improving primary health care and basic education. The broader theme that we identified as relevant for the education sector to learn from healthcare in respect of RBF use is that quality of service provision matters and hence may be a useful focus area. A great deal of attention has been paid to increasing access to education (proxied for example by the number of students enrolled in basic education), but less attention has been focused on the actual quality of teaching, which we know is strongly related to learning

outcomes. This is to some degree synonymous with poor quality service provision in healthcare, which we also know has a strong relation to health outcomes.

Concern around quality of teaching has been exacerbated by the Covid-19 pandemic and the need for improvement is now paramount. Quality of teaching is particularly critical where students need basic (literacy, language and mathematics) and transferable skills including critical thinking and problem solving, in order to advance their futures. Outcomes in education and healthcare may take very different forms but it is clear that the quality of service provision is critical to improvement in both. We note that in some specific instances, healthcare may rely more heavily on inputs (medicines for example) than might be the case in education, but this does not downplay the role of quality service provision in both sectors. For ease of reference and to cover all bases of interest we have mapped out at the start of this document what similar inputs, activities, outputs and outcomes may look like between health and education, and used this to draw out our learnings from the literature.

With the above in mind, there are many useful analogies that can be drawn between the more specific focus areas for RBF in health and education, respectively. The education literature review shows that, under certain conditions, educational gains may be advanced through the use of intervention designs using incentives to teachers. In the complex designs employed in education, external agents are generally incentivised to manage the program, which includes such functions as providing books to schools. Thus, grants to institutions do not frequently feature in such programs, although we do discuss two examples of interventions involving school grants below. In addition, the education literature reveals that intervention designs which incorporate a pedagogical element are widely reported as resulting in learning gains, outcomes for which all other intervention designs have produced mixed results.[96] We discuss intervention types under these broad categories below, drawing on the health and education literature in parallel.

6.1 Intervention types

Incentives to workers

Incentivising teachers may have an analogy to incentives to health workers, where these are most commonly in the health literature remunerated with payments to facilities and then there are usually specific rules about how much of what is attained at the facility level can be paid to healthcare workers in the form of bonuses, and how much can go into physical items for the facility. These are often centred on improving areas of health such as maternal and new-born health, HIV, reproductive health and basic vaccination coverage. We have looked at many programs that incentivise healthcare workers, across Mozambique, Burundi, Tanzania, Cameroon, Zimbabwe, and many other LMICs. These programs in health usually target activities such as complete servicing (carrying out a healthcare check completely), outputs such as service volumes (how many checks, visits, vaccinations were carried out), and outcomes such as reductions in neonatal deaths and improved child health.

The Foundation Assisted Schools (FAS) program in Pakistan is the closest in design to the health programs described in the previous paragraph, although, strictly speaking, this example does not fit under the heading of ‘incentives to workers’. In this case the government is using a non-profit body to fund low-cost

independent schools, under certain conditions. Schools are funded, and thereby teachers benefit by obtaining employment. Effectively, the government has passed on the management of a significant number of schools serving poor families to the schools, with grants conditional on meeting enrolment and performance targets. The schools established in this way in Pakistan are part of a group of low-cost private schools, supported by the state or through fees from parents, which boost enrolment significantly in a number of LMICs.[69] This is a model worth further investigation from an RBF perspective, particularly under conditions of low or skewed (low participation by girls) enrolment, or poor civil service accountability. However, the question must be asked: would not be better, from a long-term, systemic perspective, to build capacity in the public sector, than to rely on the private sector?

Often multiple types of items are of focus within one healthcare program, sometimes with both quantity (for example, how many consults did mothers get through their pregnancy) and quality (for example, was the consult thorough and complete) components. A lesser volume of RBF studies in healthcare focus on inputs such as medicines. Nevertheless, we have included these, such as the program in Tanzania that provides many learnings about incentivising the avoidance of medicine stock-outs (which may be useful in respect of learnings for book value chains in education).

Such RBF programs in the healthcare sector hold important lessons for education. Most notably, paying institutions to achieve improved outcomes, with a bonus to workers conditional on progress, is likely to incentivise workers to work toward achieving improved outcomes. Somewhat counter-intuitive to this expectation, analogous programs in education have found that bonuses paid to individuals lead to better performance than group bonuses, although both modalities lead to positive results.[77] This is an issue for further investigation.

However, in education the more frequent approach to worker incentives is to pay the bonuses directly to teachers, conditional on improved attendance and/or learning outcomes. The first of these (improving attendance) is more successful than the second (learning outcomes), and programs of the latter type show mixed results. The most likely explanation for the difficulty in raising learning outcomes is that teachers lack the capacity to do so as opposed to lacking the will. There is some evidence that, where teacher incentive schemes are linked to capacity building, they are more successful, but the data is scanty. However, capacity building interventions, on their own and not utilising RBF funding, are showing encouraging positive effects.

Over and above the dearth of positive results on learning outcomes exhibited by teacher incentive schemes (P4P), there are a number of reservations concerning the advisability of such interventions, chief of which is whether they are sustainable, and whether there are unintended consequences. Care should be taken in providing teacher incentive payments to make sure that these payments can be sustained over the long term, and do not become an unsustainable fiscal burden for the government or donors. It is important for evaluations to collect information on cost and cost-effectiveness, as these factors have important implications for sustainability. The literature also cautions against emphasising extrinsic motivation. Indeed the teacher incentive program from Andhra Pradesh described above put the emphasis on eliciting teachers' intrinsic motivation, and the evaluation claims this as an important element in its success. This learning was strongly echoed in the health literature.[77]

Systemic / system support incentives

The overriding priority in education research currently is focused on finding ways of improving learning outcomes in the foundation disciplines of language, literacy and mathematics. Failure to learn these skills in the early grades leads to diminished life opportunities. Key to achieving this goal is to improve the effectiveness of teaching. Over the past two decades interventions focused on improving pedagogy have emerged as the most effective tools for teacher capacity building and raising learning outcomes in many LMICs on all continents, but to which RBF principles have not yet been applied. There seems to be no reason why RBF should not be at least as effective as conventional funding forms when incorporated into structured pedagogy designs. We note that if so applied, this would overlay or deepen an approach to RBF where teachers are incentivised.

Here analogies in healthcare may include the focus of RBF programs on stock management (for example medicines in healthcare, as mentioned above), and healthcare checklists of items tracked for incentives to ensure complete and quality service provision (which in the health literature is often part of the RBF design when incentivising health workers). More than one health study used a "quality index score" to motivate healthcare workers not only to do procedures, but to do them fully, translating their knowledge into better practice. One healthcare study (in Cameroon) mentioned coaching, but did not explain this in any detail, but did importantly indicate that feedback loops need to be in place to check for any obstacles to RBF working and timeously address them. Whilst we note distinctions in some types of activities or processes that are required under each of education and health, it is possible that future analogies for RBF in education might address items such as structured pedagogy, which combines daily lesson plans, classroom materials and regular in-class coaching.

Because of their complex, multifaceted nature, pedagogically-focused interventions in education are inclining towards adopting outcomes-based approaches, including but not limited to Impact Bonds, with non-state training and resource providers incentivised to support schools to deliver improved learning outcomes. In such designs, a range of metrics might be used to track progress and trigger payments at various levels, including inputs (such as materials, training and coaching), outputs (monitoring the use of lesson plans) and outcomes (test scores). It should be noted, however, that education Impact Bonds launched or in procurement to date in India and sub-Saharan Africa have tended to utilise only enrolment of out-of-school children and improved learning outcomes as payment triggers, intentionally leaving inputs and activities to be flexibly determined by support providers based on locally-determined needs.[97–100]

6.2 Considerations in the design of RBF programs

This work aimed to flesh out exactly what made the difference in RBF working or not working in both the health and education sectors. This led us to the following broad guidance (drawn from 28 individual peer reviewed studies and grey literature on health RBF interventions in LMICs, including two meta-analyses and four systematic reviews):

1. **What is the context?**

Section 5.1 discusses how both health and education have broader socioeconomic determinants that should be considered when implementing an RBF program. The **environment outside of the school or healthcare facility** matters when improving outcomes.[62] This should be understood, together with understanding any **broader political economic context** that may be relevant, for example are complementary programs or economic reforms in place or not, are there value chain considerations that need to be understood and that otherwise may form constraints to improvement, etc.[58] [59,60]

The **current funding available** to the education or healthcare setting should also be understood; if funding is severely lacking to begin, this may impede the success of RBF and a focus on basic inputs may need to precede a focus on service quality. Understanding the starting point is key as this may influence improvement in other ways too; areas with lesser starting points in terms of service provision may simply improve more in the face of RBF, whilst those closer to a well-performing equilibrium may have less room for improvement.[56]

If a project is dependent in any way on government participation, a key contextual consideration is the **state of the responsible bureaucracy**: level of corruption, HR processes and accountability mechanisms, national testing programs, M&E capacity and data systems. RBF projects need to cater for weaknesses in government systems and processes but, whatever the strengths and weaknesses, government buy-in is a prerequisite for success.

Core to this full understanding of the context in which a RBF program is being considered, is to **clearly articulate what the problem is and the mechanism by which RBF will help**. From this the program can be suitably designed and directed to clearly target the issues understood.

2. **What is being incentivised?**

As mentioned in section 4.2, our focus in the literature stresses the **importance of structural and process quality** as important aspects to incentivise the quality of service provision in both education and healthcare. The individual studies described in section 5.2 highlight that there can also be **feedback loops and hence dynamic effects here**; incentive payments received by healthcare facilities for performance are often allowed to be spent on both staff payments and some degree of facility inputs. This naturally will have dynamic effects over time, where structural and process quality improvements work in conjunction and gain momentum. A healthcare study in Burundi concluded that it was unclear whether results were due to more facility resources or a change in provider behaviour[88]; this finding is not uncommon in the healthcare literature, particularly when various other programs are ongoing at the same time.

Many individual studies are also highly complex in design, with many different items being incentivised all at once. Care should be taken in ensuring that **what is being incentivised is simple enough to be understood** by those individuals or institutions targeted. In some cases, **understanding the mechanisms through which quality improvements are achieved** may require both qualitative process information and quantitative data methods.

Glewwe and colleagues describe an experimental teacher incentive program rewarded Kenyan primary school teachers for increased exam scores and sanctioned them for students not taking

the exams[28]. Exam participation increased among enrolled students and scores increased only for exams linked to the incentives, but had not for those non-target exams; these effects were accompanied by an increase in test preparation sessions.

In their review of 15 teacher incentive programs implemented in developing countries and Israel Bruns et al note that, in order to operate at scale, such interventions are frequently designed around reliable testing systems which produce credible results, supported by a high-quality national data system which disseminates information to stake-holders by means of periodic school report cards, and may be linked to bonuses for teachers and/or schools[25]. In the case of Brazil, the latter report both on student learning outcomes and grade progression, thus discouraging undesired teacher behaviour of boosting learner scores by holding back weak students. Such information-for-accountability programs are intended both to strengthen parents' voice in school matters at the school council and state levels, and to inform school personnel about their schools' instructional quality and academic performance. The authors conclude that impact evaluations of such programs produce mixed results regarding the potential of information for accountability to improve learning outcomes.

In education what is being incentivised may focus on higher school enrolment, teacher attendance, or better learning outcomes, but this is an area which is not well understood, to a large extent due to the paucity of evidence. Teacher incentives can be successful, particularly when it is clear what is expected of teachers such as more regular attendance at school, although this result is open to gaming[89]. The effects of teacher incentives on learning outcomes in developing countries are decidedly mixed, and the REACH report concludes that the best way to reconcile these divergent findings for now is to recognize that design and context matter a lot[89]. For example, the introduction of an incentive scheme can make the teachers think that they are not trusted, leading to a drop in their morale: program designers therefore need to think about the realities of the implementation of a specific program beyond its theoretical design, paying due attention to the way this is conveyed to the agents themselves.

But the most likely explanation for the largely disappointing effects of teacher incentives on learning outcomes is that teachers are unable to achieve the desired outcomes because of their own poor education: they lack the knowledge and skills required for effective teachingThe education literature tells us that improved learning is best achieved through teacher capacity building, and that, in turn, is most effectively advanced through pedagogy-focused interventions, starting with the foundation disciplines in the early grades[20].

3. **Who is being incentivised?**

Ensuring that **those who are incentivised have control** over the targeted actions and/or outcomes is key. As the Tanzanian example in section 5.3 illustrates, in some instances, **those throughout the value chain** will need to be considered (e.g. district managers, facility managers, front-line workers, etc.), to ensure that they are all working towards the same goals.[72] This example focused on avoiding medicine stock-outs, where the involvement of district managers was key. It showed that one needs to be **cognisant of factors beyond one's control** such as a global shortage of ARVs, ensuring that a program is not lost due to unavoidable challenges such as these.

A healthcare program in Rwanda described in section 5.4 concluded that **more skilled or trained workers respond better to incentives.**[84] Whilst this may logically be the case and hence one should ensure that those incentivised have the necessary training, one should also be **careful to understand how and why** this phenomenon may occur to clear any biases that may misguide any remedies (focused for example on upskilling or more clearly explaining the RBF program). Core to this part of the RBF program design is to **ensure that those incentivised, measured, and paid align.** Where funds may not reach those whose actions are being incentivised, success may be limited or unsustainable. This was for example the case in a healthcare study in Rwanda that focused on health worker cooperatives, where it was found that funds may not have sufficiently reached those whose actions were targeted or that the amount received was not enough to warrant the additional effort.

Although the evidence from education is somewhat mixed as to whether incentives are most effective when paid to individuals or groups, the message from health is clear: **incentives at the institutional level may promote teamwork** (particularly where co-workers actions are observable to others), and individual bonuses are conditional on achieving institutional targets.

4. **How is it being incentivised?**

The literature described in 5.4 also widely points to the fact that the **size, frequency and relationship of payments to marginal effort all matter.**[85] [86] Where payments are delayed or too small relative to the effort required, success may again be limited. A healthcare study in Rwanda showed that **higher payments led to better results**.[84] Other programs such as that in Mozambique showed that **efforts requiring the least marginal effort will be influenced most easily,** payment level aside.[23] In the health literature payments are often at **monthly or quarterly frequency,** increasing saliency of programs. In the calculation of what is paid out, many programs have both a **quantity and quality component**, such that the total payment is calculated as a weighted sum of the number of provided services multiplied by their unit payment, multiplied by a quality bonus, where the latter had a range depending on a score obtained from a checklist assessment of quality of service provision.

An important lesson from the health sector here is that **payments should be put in the hands of those closest to the services provided** and that the **users of those funds should enjoy a degree of autonomy** in disbursing them.[65] Lessons from the health sector conclude that some degree of autonomy in the hands of the health service provider tends to optimise the utilisation of these funds. In the health literature **this degree of autonomy may still have some structure,** for example, where a portion of the funds achieved by the facility are allowed to be allocated to staff, and another portion to structural inputs. In the latter, facility managers may have the ability to use funds to hire additional staff or procure inputs that they know are needed at their facility to support their workers. Importantly, differences in initial starting points where some facilities may be lacking basic inputs are overcome by **equity weightings, where lower-resourced areas get higher incentives**.

The health literature indicates that it may be better to **target continuous actions rather than thresholds** (meaning that incentive payments are made for each incremental action starting from

zero, rather than triggering an incentive payment only when one reaches a target of for example x healthcare checks in a portion of time, where this may lead to declined motivation after that target, or in the face of one not being able to reach that target at all). Similarly, incentives should be **absolute to an individual or institution rather than relative to others**. Where the latter is employed to incentivise competition, one should be careful of any unintended consequences and reduced motivation, especially with different starting points (for example, equity considerations in more or less well-resourced areas). Loss aversion is also pertinent in RBF; **rewards for good performance** will be better received than penalties for poor performance. This dovetails well with a final point from the literature on the basis for payment, which is that individual incentives may be better received when they are **linked to bonuses, rather than basic salary payments.** In education, where learning outcomes are targeted, benchmarks drawn from research findings may provide realistic targets, which could be supplemented by offering a bonus for achievements which exceed the target benchmark.

5. **How is it being supported?**

The health literature described in section 5.5 highlights that RBF programs may benefit from receiving support to achieve RBF aims. With some degree of autonomy for facility managers, **support in decision making** may be needed, at least to begin. Also, where challenges may arise, feedback from figures such as district managers may need to be rapid, to ensure **bottlenecks are identified and cleared**. **Checklists** of what needs to be done to achieve RBF are often relied on in the health literature to provide guidance. [79[65]

The sustainability of these interventions is also an important consideration. Given that most RBF interventions are donor supported through funding and technical assistance, the **successful handover of these programmes is essential to ensure their long-term success**.[101,102] As part of a successful handover process, it is important to ensure local buy-in for and ownership of the RBF programme within government as part of the implementation process as early as possible.[101–103] In addition to this, clear guidelines need to be drafted to guide how the transfer of ownership of the RBF program from donor to government should happen, which in the case of Cameroon may even include the drafting of supporting legislation, and this transition needs to be owned and planned by both parties.[103] A small number of successful interventions in education have been taken to scale. None of these programs utilise RBF modalities but, given their success in raising learning outcomes, hold significant potential for the expansion of RBF designs. In a study of 8 such programs Stern et al noted that six could be characterized as structured pedagogy (Appendix 1) programs, which typically had structured learning materials for students matched with teachers' guides for teachers, focused training on new instructional methods for teachers, and teacher support systems that included coaching[104]. One program was based on a TaRL design (Appendix 1), while the eighth also had a primary focus on early literacy and numeracy but, in addition, included components on building leadership capacity at school and district levels.

Furthermore, the study concluded, programs were able to achieve success when they were seen as important priorities by education system officials, a step which is likely to enhance the chances of being sustainable. Having evidence to demonstrate the success of the particular instructional model being introduced was one way to convince ministry counterparts of the value of the

program. Several programs systematically developed government capacity and transferred responsibility to government structures in stages. Ultimately, the key ingredient for producing measurable improvements in student learning outcomes was the change in organizational culture which the programs managed to engender in government.

6. **What are possible unintended consequences?**

It would be incomplete to consider the impact of an intervention without accounting for the potential unintended consequences. Unintended consequences should be carefully considered, with **feedback received in a timely manner and acted on**. Section 5.6 describes how **independent and frequent auditing** is crucial. Some schemes impose large **penalties for inaccurate reporting** (with cognisance that RBF in itself may simply improve reporting, as was mentioned in one Mozambican healthcare study).[23]

Equity considerations should also be made, taking into account potential bonuses for lower-resourced areas or institutions, and also **considering whether there are any demand-side impediments** where fees act as barriers (as mentioned, demand-side payments have mainly been shown to be relevant in these instances for the poor). Finally, in one Zimbabwean healthcare study an unintended consequence of RBF was increased patient load; again **structural quality** in the form of sufficient staff may be a necessary consideration ahead of RBF implementation focused on quality of service provision.

Another unintended consequence might be that a RBF program and related study does not have **external validity**. Our literature review has highlighted that a study needs to be carried out over a long enough period of time to ensure external validity (up to 18 months was recommended by an Mozambican healthcare study,[23] in order to **allow adjustment to happen and hence results to materialise**), it should also be considered whether the incentives will **remain in place or be adjusted** over time, and whether they are **desirable and sustainable**. Studies from our literature review also highlight that individuals and institutions often require **time to be fully informed of how a RBF program works, time to adapt and reach a new way of working.** Programs and studies that are carried out over a short period may hence lack findings on both external and internal validity, time to adjust and understand is essential, as is **sufficient communication, collaboration and feedback** on the working of the program and any bottlenecks.

7. **How is it being evaluated?**

Finally in section 5.7 describes how untangling the respective contributions to success or failure of the various elements of complex programs requires a sophisticated M&E system, including independent evaluations of both programme results and, ideally, processes. There are numerous examples of evaluations, in both health and education, which end inconclusively because they were commissioned too late due to inadequate planning, or were not sufficiently robust in design, with shortcomings sometimes necessitated by insufficient budgets. **Rigorous evaluations** are expensive, but the failure to advance knowledge of program impact inhibits progress in designing effective and efficient interventions. We have mentioned above that **context is key**, and in the

absence of this, determining whether RBF did or did not work and why may be challenging. In some cases RBF may not show results because certain individuals were not involved, or because of less room for improvement. **Understanding the detail reveals insights** into how to improve and best target RBF.[105]

As stated, the above guidance around what contributes to RBF working in healthcare is drawn both from looking at results between and within studies, and the reasons for these. To ensure that quality improvements are sustained, a focus on governance, management structures, and financial autonomy, as well as the inclusion of the perspectives and experiences of patients and health workers is essential.[72] In education, these considerations manifest in complex, multi-faceted intervention designs, generally managed by outside agents who are incentivised to achieve learning outcomes. Nevertheless, cognisance should be had for the fact that every study context, as highlighted above, will be vastly different, and caution must be exercised here.

References

1. The World Bank. Results-Based Financing (RBF) and Results in Education for All Children (REACH) [Internet]. World Bank [cited 2022 Jan 13];Available from: https://www.worldbank.org/en/programs/reach

2. The World Bank. Banking on Impact: What You Need to Know about Results-Based Financing [Internet]. World Bank [cited 2022 Jan 13];Available from: https://www.worldbank.org/en/news/feature/2019/06/28/banking-on-impact-what-you-need-to-know-about-results-based-financing

3. United Nations. THE 17 GOALS | Sustainable Development [Internet]. [cited 2021 Oct 12];Available from: https://sdgs.un.org/goals

4. Lockheed M, Prokic-Bruer T, Shadrova A. The Experience of Middle-Income Countries Participating in PISA 2000-2015 [Internet]. OECD; 2015 [cited 2021 Oct 13]. Available from: https://www.oecd-ilibrary.org/education/the-experience-of-middle-income-countries-participating-in-pisa-2000-2015_9789264246195-en

5. World Bank Group. Learning Poverty [Internet]. Learning Poverty2019 [cited 2021 Oct 12];Available from: https://www.worldbank.org/en/topic/education/brief/learning-poverty

6. World Bank Group. What is learning poverty? [Internet]. World Bank2021 [cited 2021 Oct 12];Available from: https://www.worldbank.org/en/topic/education/brief/what-is-learning-poverty

7. Mawoyo M, Vally Z. Improving Education Outcomes in Low- and Middle-Income Countries: Outcomes-Based Contracting and Early Grade Literacy. Journal of Learning for Development 2020;7(3):334–48.

8. Besharati N, Fleisch B, Tsotsotso K. Interventions to improve learner achievement in South African. A systematic meta-analysis [Internet]. In: Systematic Reviews of Research in Basic Education in South Africa. Stellenbosch: African Sun Media; 2021 [cited 2021 Oct 13]. Available from: https://africansunmedia.store.it.si/book/systematic-reviews-of-research-in-basic-education-in-south-africa/1158608/

9. UNESCO. Global Education Monitoring (GEM) Report [Internet]. 2020 [cited 2021 Oct 12]. Available from: https://en.unesco.org/news/global-education-monitoring-gem-report-2020

10. Glewwe P, Hanushek E, Humpage S, Ravina R. School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010 [Internet]. Cambridge, MA: National Bureau of Economic Research; 2011 [cited 2021 Oct 12]. Available from: http://www.nber.org/papers/w17554.pdf

11. Bold T, Filmer D, Molina E, Svensson J. The Lost Human Capital: Teacher Knowledge and Student Achievement in Africa [Internet]. World Bank, Washington, DC; 2019 [cited 2021 Oct 12]. Available from: http://hdl.handle.net/10986/31673

12. Crawfurd L, Pugatch T. Teacher Labor Markets in Developing Countries [Internet]. Rochester, NY: Social Science Research Network; 2020 [cited 2021 Oct 12]. Available from: https://papers.ssrn.com/abstract=3542654

13. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. The Lancet Global Health 2018;6(11):e1196–252.

14. Castro C de M, Musgrove P. Why education and health are more different than alike. Ensaio: aval.pol.públEduc 2016;24(91):477–91.

15. Paul E, Albert L, Bisala BN, Bodson O, Bonnet E, Bossyns P, et al. Performance-based financing in low-income and middle-income countries: isn't it time for a rethink? BMJ Glob Health 2018;3(1):e000664.

16. King JJC, Powell-Jackson T, Makungu C, Hargreaves J, Goodman C. How much healthcare is wasted? A cross-sectional study of outpatient overprovision in private-for-profit and faith-based health facilities in Tanzania. Health Policy and Planning 2021;36(5):695–706.

17. Sun X, Liu X, Sun Q, Yip W, Wagstaff A, Meng Q. The Impact of a Pay-for-Performance Scheme on Prescription Quality in Rural China: An Impact Evaluation. :33.

18. UNESCO Institute for Statistics. New Methodology Shows that 258 Million Children, Adolescents and Youth Are Out of School. USIS Fact Sheet No 56. 2019;

19. Besharati N, Fleisch B, Tsotsotso K. Interventions to improve learner achievement in South African. A systematic meta-analysis. In Maringa, F. (Ed): Systematic Reviews of Research in Basic Education in South Africa. [Internet]. Stellenbosch: African Sun Media; 2021. Available from: https://africansunmedia.store.it.si/book/systematic-reviews-of-research-in-basic-education-in-south-africa/1158608/

20. Snilstveit B, Stevenson J, Phillips D, Vojtkova M, Gallagher E, Schmidt E, et al. Interventions for improving learning outcomes and access to education in low- and middle- income countries: a systematic review. 2015;

21. Lee J, Medina O. Results based financing: Learning from what works. World Bank Group;

22. Banerjee A, Banerji R, Berry J, Duflo E, Kannan H, Mukerji S, et al. From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application. Journal of Economic Perspectives 2017;31(4):73–102.

23. Rajkotia Y, Zang O, Nguimkeu P, Gergen J, Djurovic I, Vaz P, et al. The effect of a performance-based financing program on HIV and maternal/child health services in Mozambique - An impact evaluation. Health Policy and Planning 2017;32:1386–96.

24. Deaton A, Cartwright N. Understanding and misunderstanding randomised controlled trials. [Internet]. 2016;Available from: https://www.nber.org/system/files/working_papers/w22595/w22595.pdf

25. Bruns B, Filmer D, Patrinos H. Making Schools Work: New Evidence on Accountability Reforms. Washington DC: The World Bank; 2011.

26. Tooley J. Is private education good for the Poor? [Internet]. 2005;Available from: https://thefilter.blogs.com/thefilter/files/tooley.pdf

27. Anwar M, Khizar A, Haq R. An Analysis of Foundation Assisted Schools Program of Punjab as a Mechanism Influencing Pupil Cohort An Analysis of FASs Program of PEF as a Mechanism Influencing Pupil Cohort 2. 2018;1–12.

28. Glewwe P, Ilias N, Kremer M. Teacher Incentives. American Economic Journal: Applied Economics 2010;2(3):205–27.

29. Varthana raises $3 mn from Dell Foundation to expand school program | Business Standard News [Internet]. [cited 2021 Oct 12];Available from: https://www.business-standard.com/article/companies/varthana-raises-3-mn-from-dell-foundation-to-expand-school-program-117101200440_1.html

30. Impact bonds [Internet]. The Government Outcomes Lab [cited 2021 Nov 2];Available from: https://golab.bsg.ox.ac.uk/the-basics/impact-bonds/

31. Morrow S. Qualitative Research in Counseling Psychology: Conceptual Foundations. The Counselling Psychologist [Internet] 2007 [cited 2021 Nov 2];35(2). Available from: https://journals.sagepub.com/doi/10.1177/0011000006286990

32. Mission | RBF Health [Internet]. [cited 2021 Oct 20];Available from: https://www.rbfhealth.org/mission

33. Completed Impact Evaluations and Emerging Lessons from the Health Results Innovation Trust Fund Learning Portfolio | RBF Health [Internet]. [cited 2021 Oct 15];Available from: https://www.rbfhealth.org/resource/completed-impact-evaluations-and-emerging-lessons-health-results-innovation-trust-fund

34. Social Finance. Cameroon Kangaroo Mother Care Development Impact Bond 2018-2021: End of programme report [Internet]. 2021 [cited 2021 Oct 28];Available from: http://www.socialfinance.org.uk

35. Marmot, M, Wilkinson, R. Social Determinants of Health. 2nd ed. 2006.

36. Correa JA, Gutiérrez P, Lorca M, Morales R, Parro F. The persistent effect of socioeconomic status on education and labor market outcomes: Evidence from Chile's administrative records. Applied Economic Analysis 2019;27(79):62–90.

37. Thomson S. Achievement at school and socioeconomic background—an educational perspective. npj Science Learn 2018;3(1):5, s41539-018-0022–0.

38. Murtin F. Improving Education Quality in South Africa [Internet]. Paris: OECD; 2013 [cited 2021 Oct 13]. Available from: https://www.oecd-ilibrary.org/economics/improving-education-quality-in-south-africa_5k452klfn9ls-en

39. Lemiere C, Torsvik G, Maestad O, Herbst CH, Leonard KL. Evaluating the impact of results-based financing on health worker performance. The World Bank Group; 2013.

40. Curristine T, Lonti Z, Joumard I. Improving Public Sector Efficiency: Challenges and Opportunities. OECD JOURNAL ON BUDGETING 2007;7(1):42.

41. Lefafa N. Hospitals vandalised over unpaid performance bonuses [Internet]. Health-e News2020 [cited 2021 Oct 26];Available from: https://health-e.org.za/2020/12/30/north-west-hospitals-vandalised-over-unpaid-performance-bonuses/

42. Nxumalo N, Goudge J, Gilson L, Eyles J. Performance management in times of change: experiences of implementing a performance assessment system in a district in South Africa. International Journal for Equity in Health 2018;17(1):141.

43. Lohmann J, Houlfort N, De Allegri M. Crowding out or no crowding out? A Self-Determination Theory approach to health worker motivation in performance-based financing. Social Science & Medicine 2016;169:1–8.

44. Kaffenberger M, Spivak M. System Coherence for Learning: Applications of the RISE Education Systems Framework. [Internet]. 2022;Available from: https://doi.org/10.35489/BSG-RISEWP_2022/086

45. Custers T, Hurley J, Klazinga NS, Brown AD. Selecting effective incentive structures in health care: A decision framework to support health care purchasers in finding the right incentives to drive performance. BMC Health Services Research 2008;8(1):66.

46. Manongi R, Mushi D, Kessy J, Salome S, Njau B. Does training on performance based financing make a difference in performance and quality of health care delivery? Health care provider's perspective in Rungwe Tanzania. BMC Health Services Research 2014;14(1):154.

47. Cashin C, Bloom D, Sparkes S, Barroy H, Kutzin J, Dougherty S O'. Aligning Public Financial Management and Health Financing | Results for Development [Internet]. World Health Organization; 2017 [cited 2021 Oct 15]. Available from: https://r4d.org/resources/aligning-public-financial-management-health-financing/

48. Smith A, Ranchod S, Strugnell D, Wishnia J. Human resources for health planning and National Health Insurance: the urgency and the opportunity. 2018. page 23.

49. Zwisler G, Simpson E, Moodley M. Treatment of diarrhea in young children: results from surveys on the perception and use of oral rehydration solutions, antibiotics, and other therapies in India and Kenya. J Glob Health 2013;3(1):010403.

50. Roberts E, Bornstein MH, Slater AM, Barrett J. Early cognitive development and parental education. Infant and Child Development 1999;8(1):49–62.

51. Dubow EF, Boxer P, Huesmann LR. Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations. Merrill Palmer Q (Wayne State Univ Press) 2009;55(3):224–49.

52. OECD. Results base financing: Key take-aways. 2015.

53. Tovey D. How to clarify a clinical question [Internet]. BMJ Best Practice [cited 2021 Oct 12];Available from: https://bestpractice.bmj.com/info/toolkit/learn-ebm/how-to-clarify-a-clinical-question/

54. Das A, Gopalan SS, Chandramohan D. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. BMC Public Health 2016;16(1):321.

55. Kovacs RJ, Powell-Jackson T, Kristensen SR, Singh N, Borghi J. How are pay-for-performance schemes in healthcare designed in low- and middle-income countries? Typology and systematic literature review. BMC Health Services Research 2020;20(1):291.

56. Paul E, Sossouhounto N, Eclou DS. Local stakeholders' perceptions about the introduction of performance-based financing in Benin: a case study in two health districts. Int J Health Policy Manag 2014;3(4):207–14.

57. World Development Report 2018: Learning to Realize Education's Promise [Internet]. World Bank [cited 2021 Oct 13];Available from: https://www.worldbank.org/en/publication/wdr2018

58. Soucat A, Dale E, Mathauer I, Kutzin J. Pay-for-Performance Debate: Not Seeing the Forest for the Trees. Health Systems & Reform 2017;3(2):74–9.

59. Wishnia J, Goudge J. Impact of financial management centralisation in a health system under austerity: a qualitative study from South Africa. BMJ Glob Health 2020;5(10):e003524.

60. Gautier L, De Allegri M, Ridde V. How is the discourse of performance-based financing shaped at the global level? A poststructural analysis. Globalization and Health 2019;15(1):6.

61. Duran D, Bauhoff S, Berman P, Gaudet T, Konan C, Ozaltin E, et al. The role of health system context in the design and implementation of performance-based financing: evidence from Cote d'Ivoire. BMJ Global Health 2020;5(9):e002934.

62. Bruns B, Schneider BR. Managing the Politics of Quality Reforms in Education: Policy Lessons from Global Experience [Internet]. The Education Commission; 2016. Available from: http://report.educationcommission.org/wp-content/uploads/2017/01/Managing-the-Politics-of-Quality-Reforms.pdf

63. Hickey S, Hossain N. The Problem of Education Quality in Developing Countries. In: The Politics of Education in Developing Countries: From Schooling to Learning. Oxford University Press; 2019.

64. Bano M, Oberoi Z. Embedding Innovation in State Systems: Lessons from Pratham in India [Internet]. Research on Improving Systems of Education (RISE); 2020. Available from: https://doi.org/10.35489/BSG-RISE-WP_2020/058

65. de Walque D, Robyn PJ, Saidou H, Sorgho G, Steenland M. Looking into the performance0based financing black box: evidence from an impact evaluation in the health sector in Cameroon. World Bank Group; 2017.

66. OECD. Better Ways to Pay for Health Care [Internet]. OECD; 2016 [cited 2022 Jan 14]. Available from: https://www.oecd-ilibrary.org/social-issues-migration-health/better-ways-to-pay-for-health-care_9789264258211-en

67. Kovacs RJ, Powell-Jackson T, Kristensen SR, Singh N, Borghi J. How are pay-for-performance schemes in healthcare designed in low- and middle-income countries? Typology and systematic literature review. BMC Health Services Research 2020;20(1):291.

68. World Health Organization. Everybody's business: strengthening health systems to improve health outcomes : WHO's frmaework for action. Geneva: World Health Organization; 2007.

69. Tooley J. Is private education good for the Poor? [Internet]. 2005. Available from: https://thefilter.blogs.com/thefilter/files/tooley.pdf

70. Ogbuoji KKM Kenneth Munge, Gilbert Kokwaro, and Osondu. Reducing Kenya's health system dependence on donors [Internet]. Brookings2021 [cited 2022 Jan 12];Available from: https://www.brookings.edu/blog/future-development/2021/03/02/reducing-kenyas-health-system-dependence-on-donors/

71. Ogundeji YK, Sheldon TA, Maynard A. A reporting framework for describing and a typology for categorizing and analyzing the designs of health care pay for performance schemes. BMC Health Serv Res 2018;18(1):686.

72. Binyaruka P, Borghi J. Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania. Trop Med Int Health 2017;22(1):92–102.

73. Lee J, Medina O. Results based financing: Learning from what works. World Bank Group;

74. Shwartz M, Restuccia JD, Rosen AK. Composite Measures of Health Care Provider Performance: A Description of Approaches. Milbank Q 2015;93(4):788–825.

75. Makuluni R, Stones W. Impact of a maternal and newborn health results-based financing intervention (RBF4MNH) on stillbirth: a cross-sectional comparison in four districts in Malawi. BMC Pregnancy and Childbirth 2021;21(1):417.

76. Shapira G, Kalisa I, Condo J, Humuza J, Mugeni C, Nkunda D, et al. Effects of Performance Incentives for Community Health Worker Cooperatives in Rwanda [Internet]. Washington, DC: World Bank; 2017 [cited 2021 Oct 15]. Available from: https://openknowledge.worldbank.org/handle/10986/26752

77. Muralidharan K, Sundararaman V. Teacher Performance Pay: Experimental Evidence from India. Journal of Political Economy 2011;119(1):39–77.

78. Teacher pay for performance: Does it really work? [Internet]. [cited 2021 Nov 2];Available from: https://blogs.worldbank.org/education/teacher-pay-performance-does-it-really-work

79. Takian A, Rashidian A, Doshmangir L. The experience of purchaser–provider split in the implementation of family physician and rural health insurance in Iran: an institutional approach. Health Policy and Planning 2015;30(10):1261–71.

80. Mbau R, Barasa E, Munge K, Mulupi S, Nguhiu PK, Chuma J. A critical analysis of health care purchasing arrangements in Kenya: A case study of the county departments of health. The International Journal of Health Planning and Management 2018;33(4):1159–77.

81. Ogbuabor DC, Onwujekwe OE. Scaling-up strategic purchasing: analysis of health system governance imperatives for strategic purchasing in a free maternal and child healthcare programme in Enugu State, Nigeria. BMC Health Services Research 2018;18(1):245.

82. Ergo A, Paina L. Verification in performance-based incentive schemes. USAID; 2012.

83. Lavadenz F, Schwab N, Straatman H. Public, decentralized and community health networks in Bolivia. Rev Panam Salud Publica 2001;9(3):182–9.

84. Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. The Lancet 2011;377(9775):1421–8.

85. Chung S, Palaniappan L, Wong E, Rubin H, Luft H. Does the Frequency of Pay-for-Performance Payment Matter?—Experience from a Randomized Trial. Health Serv Res 2010;45(2):553–64.

86. Eijkenaar F. Key issues in the design of pay for performance programs. The European Journal of Health Economics 2013;14(1):117–31.

87. Seddon ME, Marshall MN, Campbell SM, Roland MO. Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand. Qual Health Care 2001;10(3):152–8.

88. Bonfrer I, Poel E, Doorslaer E. The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. Social Science & Medicine 2014;123.

89. Lee J, Medina O. Results based financing: Learning from what works.

90. John W Peabody, Riti Shimkhada, Stella Quimbo, Orville Solon, Xylee Javier, Charles McCulloch. Impact of performance incentives on child health outcomes: results from a cluster randomized controlled trial in the Philippines. Health Policy and Planning 2014;29:615–21.

91. Piper B, Destefano J, Kinyanjui EM, Ong'ele S. Scaling up successfully: Lessons from Kenya's Tusome national literacy program. J Educ Change 2018;19(3):293–321.

92. Argentina - Stand-Alone Impact Evaluation | RBF Health [Internet]. [cited 2021 Nov 2];Available from: https://www.rbfhealth.org/project/argentina-stand-alone-impact-evaluation

93. Antony M, Bertone MP, Barthes O. Exploring implementation practices in results-based financing: the case of the verification in Benin. BMC Health Services Research 2017;17(1):204.

94. Turcotte-Tremblay A-M, Gali Gali IA, Ridde V. An Exploration of the Unintended Consequences of Performance-Based Financing in 6 Primary Healthcare Facilities in Burkina Faso. International Journal of Health Policy and Management [Internet] 2020 [cited 2022 Jan 13];0. Available from: https://www.ijhpm.com/article_3836.html

95. Cambridge Education. Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia [Internet]. 2016 [cited 2021 Nov 2]. Available from: https://reliefweb.int/report/ethiopia/evaluation-pilot-project-results-based-aid-education-sector-ethiopia

96. Snilstveit B, Stevenson J, Phillips D, Vojtkova M, Gallagher E, Schmidt T, et al. Interventions for improving learning outcomes and access to education in low- and middleincome countries. International Initiative for Impact Evaluation (3ie);

97. Educate Girls [Internet]. The Government Outcomes Lab [cited 2021 Nov 2];Available from: https://golab.bsg.ox.ac.uk/knowledge-bank/case-studies/educate-girls/

98. Quality Education India Development Impact Bond [Internet]. The Government Outcomes Lab [cited 2021 Nov 2];Available from: https://golab.bsg.ox.ac.uk/knowledge-bank/case-studies/quality-education-india-dib/

99. Education Outcomes Fund - Sierra Leone [Internet]. The Government Outcomes Lab [cited 2021 Nov 2];Available from: https://golab.bsg.ox.ac.uk/knowledge-bank/indigo/fund-directory/INDIGO-FUND-0019/

100. Education Outcomes Fund - Ghana [Internet]. The Government Outcomes Lab [cited 2021 Nov 2];Available from: https://golab.bsg.ox.ac.uk/knowledge-bank/indigo/fund-directory/INDIGO-FUND-0018/

101. Gautier L, Ridde V. Health financing policies in Sub-Saharan Africa: government ownership or donors' influence? A scoping review of policymaking processes. Global Health Research and Policy 2017;2(1):23.

102. Wilhelm DJ, Brenner S, Muula AS, De Allegri M. A qualitative study assessing the acceptability and adoption of implementing a results based financing intervention to improve maternal and neonatal health in Malawi. BMC Health Serv Res 2016;16(1):398.

103. Sieleunou I, Turcotte-Tremblay A-M, Yumo HA, Kouokam E, Fotso J-CT, Tamga DM, et al. Transferring the Purchasing Role from International to National Organizations During the Scale-Up Phase of Performance-Based Financing in Cameroon. Health Systems & Reform 2017;3(2):91–104.

104. Stern J, Jukes M, Piper B, DeStefano J, Mejia J, Dubeck P, et al. Learning at Scale: Interim Report. 2021;

105. Spaull DN. A REPORT TO PRESIDENT RAMAPHOSA & MINISTER MBOWENI. :21.

106. Webster J, Kayentao K, Bruce J, Diawara SI, Abathina A, Haiballa AA, et al. Prevention of malaria in pregnancy with intermittent preventive treatment and insecticide treated nets in Mali: a quantitative health systems effectiveness analysis. PLoS One 2013;8(6):e67520.

107. Bulat J, King S, Spratt J. Evaluation of Mali's mother-tongue early grade "Read Learn Lead" program: Endline Report. 2013 [cited 2021 Nov 2];Available from: https://www.rti.org/publication/evaluation-malis-mother-tongue-early-grade-read-learn-lead-program-0

108. Bruns B, Filmer D, Patrinos H. Making Schools Work through Accountability Reforms. 2011. page 211–51.

109. Mizala A, Schneider BR. Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990–2010). Governance 2014;27(1):87–109.

110. Wales J, Magee A, Nicolai S. How does political context shape education reforms and their success? Lessons from the Development Progress project. ODI; 2016.

111. Fleisch B. Mona Mourshed, Chinezi Chijioke and Michael Barber: How the world's most improved school systems keep getting better. Journal of Educational Change 2011;12.

112. Darling-Hammond L, Hyler M, Gardner M. Effective Teacher Professional Development [Internet]. Learning Policy Institute; 2017 [cited 2021 Nov 2]. Available from: https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report

113. Schleicher A. Educating Learners for Their Future, Not Our Past. ECNU Review of Education 2018;1(1):58–75.

114. Liu W-C, Koh C, Chua BL. Developing Thinking Teachers Through Learning Portfolios [Internet]. In: Tan O-S, Liu W-C, Low E-L, editors. Teacher Education in the 21st Century: Singapore's Evolution and Innovation. Singapore: Springer; 2017 [cited 2021 Nov 2]. page 173–92.Available from: https://doi.org/10.1007/978-981-10-3386-5_10

Appendix 1: Examples of proven pedagogical approaches that have improved outcomes in education

Structured pedagogy

Typically, structured pedagogy programs provide structured lesson plans for teachers, supported by on-site coaching and a supply of learning materials for students; generally, they provide, at most, only moderate amounts of teacher training. The meta-analysis conducted by Snilstveit et on the effects of structured pedagogy interventions on Language arts test scores drew on 67 separate effect sizes from 18 different studies. The overall average effect of pedagogy interventions on language arts test scores was 0.23 SD, ranging from -0.14 to 0.90. In education, effect sizes in the order of 0.2-0.3 SD are generally accepted as significant. The effects of SPs in maths test scores were reflected in 24 separate effect sizes from 14 different studies, overall average effect size of 0.14, ranging from -0.09 to 0.74.[96]

Qualitative studies reveal that programs which adopt a structured pedagogy design frequently suffer from problems in implementation, such as failure to deliver tools and supplies.[106,107] For example, in the case of the Read-Learn-Lead (RLL) intervention in Mali, schools did not receive the planned materials and teacher training. The evaluators also note that the intervention started later than planned due to an overly ambitious plan for implementation, that teacher turnover caused problems and that classes were not held as often as intended. There is evidence in some programs of this type, such as the PRIMAR intervention in Kenya, that teachers were not knowledgeable or experienced enough to fully understand their training.[91]

The Snilstveit et al (2015) meta-analysis notes that most of the evaluated programs in their sample were either experimental or small scale/pilot interventions.[96] Only two of the programs were nationally-implemented at scale. In Chile, 90 percent of all primary and secondary schools were eligible to participate in the SNED program). In Mexico, 51 to 60 percent of all basic education teachers participated in the Carrera Magisterial at some point between 1991 and 2002.

With regard to small-scale experimental programs, a question has to be asked concerning the value of the evidence they provide. This question was posed by Bruns, Filmer and Patrinos (2011), when they wondered to what extent experimental studies are likely to be implemented in a nonexperimental context or at scale.[108] This question points to the distinction between internal validity (a program is effective under carefully controlled experimental conditions) and external validity (the program is effective under conditions of messy reality). According to Bruns et al (2011: 24):

"Evaluation results from a program that has been carefully monitored and "tended" but has little resemblance to what could be implemented on a large scale may also be less relevant than evaluations of programs implemented systemwide."[108]

One program employing a structured pedagogy design which has passed the external validity test, subsequent to the publication of the Snilstveit et al meta-analysis, is the Tusome intervention in Kenya. Following the success of a pilot, Tusome was successfully implemented in all primary schools in the country, resulting in a marked improvement in a range of indicators: letter/sound fluency, oral reading fluency, comprehension and emergent reading.[91] The success of Tusome is attributed to the involvement of Department of Education officials in the implementation of the program, from inception to systems-level rollout. Evaluation results indicated that several of the key elements for successful scale-up and take-up by government were put in place, including setting benchmarks for Kiswahili and English learning outcomes, and communicating these all the way down to the school level; developing functional but simple accountability and feedback mechanisms to track performance against benchmark expectations; and using feedback data to provide greater levels of instructional support within Kenya's county level structures for education quality support.[91] This relates to the question posed below under the health literature of what unintended consequences (such as a lack of external validity) might evolve and how to pre-empt them, as well as how to support RBF programs.

Teaching at the Right Level (TaRL)

The Teaching at the Right Level (TaRL) approach is another pedagogy-focused initiative which has achieved success in raising test scores in early grade reading and maths. Like the structured pedagogy approach, the model also involves teacher training and coaching and the supply of materials. In addition, children are grouped according to their level of learning by means of baseline tests, taken out of their grade-level classes for an hour a day, and taught according to their learning levels.[22] Not only has the effectiveness of the model been demonstrated in numerous 'proof of concept' experiments on a small scale, but a number of studies detail the complex research and development efforts, over a period of more than a decade, required to establish that it is scalable and implementable by government, eventually reaching more than 33 million children in India. This links with what has been discussed under the health literature around what is being incentivised.

Education reform in Chile

Over the past three decades, Chile has shown impressive economic performance, outpacing its neighbours in terms of growth rates while maintaining economic stability. In terms of international test scores, Chile moved from being on a par with its larger Latin American neighbours in 2000 to being the leader in all subject areas since 2003.[57,109,110] Chile, along with two OECD members, were the only countries to improve their PISA scores by more than 20 points in reading assessments over the decade 2000-2009; at the same time Chile demonstrated science improvements over this period that were above the OECD average.[57] The country has also witnessed a steady improvement in literacy rates over the last three decades.[58] By 2009 literacy was almost universal amongst youth and adult populations and the country's primary level pupil to teacher ratio fell impressively over 1999-2010 from 32 to 23.5.[57]

The World Development Report 2018 attributes the success of Chile's educational reform program to four key features of the policy-making environment in the 1990s and early 2000s: Emphasis on consensus in politics and policy; multiple, mutually reinforcing interventions to improve education quality; investment and targeting of financial resources; and teacher professionalisation and employment conditions.[57]

Regarding the last of these, to combat the low morale of Chile's teachers at the end of the Pinochet regime, the 1991 Teachers' Statute restored municipal teachers' status as public servants and reintroduced central bargaining for wages. Rising teacher morale and pay were successful in improving the quantity and quality of applicants to the profession. Increases in the number and quality of teachers were key enabling factors for a range of other reforms. The Ministry of Education introduced two waves of teacher performance incentives. In 1995, a new teacher bonus was introduced and allocated according to school SIMCE scores (The Sistema de Medición de la Calidad de la Educación, a battery of tests administered nationally). It was followed, in 2003, by the National Teacher Evaluation Programme, which concentrated specifically on increasing teachers' skills.

The Chilean case is one of the most recent examples of a category of educational reform which has largely escaped the attention of the economics of education literature. This caveat quite possibly arises because most of the countries and smaller jurisdictions falling into this category of reform commenced their programs before the advent of the international testing movement, depriving statisticians of the data required for quantitative studies. Nevertheless, such reform has attracted the attention of a number of the world's most respected scholars of education, writing on developments in certain educational jurisdictions, including Singapore, South Korea, Finland, Japan, Ontario, Boston and Shanghai.[111–113]

These are the world's high-performing systems, many of which started from modest to very modest beginnings and, over a period of 3 or 4 generations, achieved their current status. The defining characteristic is that they adopted a systemic approach, be it confined to the preparation, deployment and management of teachers as has been the case in Finland since the 1970s,[112] or whether it is part of an even more comprehensive reorganisation of the economic, administrative and social spheres, such as has been the case in Singapore since gaining independence in the late 1960s[114] and, more recently, Chile.[57] This links strongly with insights from the health literature on the need to fully understand the context in which RBF programs are implemented.

Appendix 3: Overview of Literature Reviewed

Health Literature

Author	Name of paper	Full citation	Type of incentive/ RBF	Target group	Sample size	Level in results chain where disbursement is triggered in health (compare to educ) (inputs, activities, outputs, outcomes, impact)	Measures of quality	Impact on quality

Kovacs, R.J, Powell-Jackson, T., Kristensen, S.R., Singh, N., and Borghi, J. (2020) https://bmchealthservres.biomedcentral.com/track/pdf/10.1186/s12913-020-05075-y.pdf	How are pay-for-performance schemes in healthcare designed in low- and middle-income countries? Typology and systematic literature review.	Kovacs, R.J., Powell-Jackson, T., Kristensen, S.R., Singh, N. and Borghi, J., 2020. How are pay-for-performance schemes in healthcare designed in low-and middle-income countries? Typology and systematic literature review. BMC health services research, 20(1), pp.1-14. Available: https://bmchealthservres.biomedcentral.com/track/pdf/10.1186/s12913-020-05075-y.pdf	Meta analysis (very useful overview of RBF in healthcare)	-	41 schemes	-	-	-

Oxman AD, Fretheim A. (2008) https://focusintl.com/data/documents/RBM094-Rapport_08_16_Results-based-financing.pdf	An overview of research on the effects of results-based financing	Oxman A.D. and Fretheim A., 2008. *An overview of research on the effects of results-based financing.* Report Nr 16-2008. Oslo: Nasjonalt kunnskapssenter for helsetjenesten. Available: https://focusintl.com/data/documents/RBM094-Rapport_08_16_Results-based-financing.pdf	Meta analysis of older papers	-	10 systematic reviews	-	-	-

Gertler P, Giovagnoli P, Martinez S. (2014) (Policy Res Wok Pap World Bank) Available: https://documents1.worldbank.org/curated/en/910221468002421288/pdf/WPS6884.pdf	Rewarding provider performance to enable a healthy start to life: evidence from Argentina's Plan Nacer	Gertler, P.J., Giovagnoli, P.I. and Martinez, S., 2014. *Rewarding provider performance to enable a healthy start to life: evidence from Argentina's Plan Nacer*. World Bank Policy Research Working Paper, (6884). Available: https://documents1.worldbank.org/curated/en/910221468002421288/pdf/WPS6884.pdf	Argentina's Plan Nacer provides insurance for MNCH (maternal, newborn and child health) to uninsured families. This was a supply side P4P national program based on an insurance program that allocated funding to provinces based on enrolment of	MNCH services in Argentina, payments made to provinces (with onward payment to facilities)	7 provinces - 228,656 prenatal visits, 108,535 tetanus vaccines, 282,042 cesareans, 274,078 birth weight, 274,078 low birth weight, 248,840 Apgar scores, 131,943 neonatal mortality	Activities, outputs and outcomes	Many "tracers" on first pregnancy checkup before 20 weeks, five minute apgar scores, birth weight, mother given tetanus vaccine and other tests, proper reviews in case of death, measles and other vaccine coverage, mothers receiving counselling post birth, children with complete record of preventative checkups, staff trained to provide care to indigineous population	The paper finds that the program increases the use and quality of prenatal care as measured by the number of visits and the probability of receiving a tetanus vaccine. Beneficiaries' probability of low birth-weight is estimated to be reduced by 19 percent. Beneficiaries have a 74 percent lower chance of in-hospital neonatal mortality in larger facilities and approximately half this reduction comes from preventing low birth weight and

			beneficiaries and adding incentives based on indicators of the use and quality of MNCH services (provinces use this to pay facilities to provide MNCH services to beneficiaries)					half from better postnatal care.

Rajkotia Y, Zang O, Nguimkeu P, Gergen J, Djurovic I, Vaz P, et al. (2017) (Health Policy Plan) Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5886140/	The effect of a performance-based financing program on HIV and maternal/child health services in Mozambique - an impact evaluation	Rajkotia, Y., Zang, O., Nguimkeu, P., Gergen, J., Djurovic, I., Vaz, P., Mbofana, F. and Jobarteh, K., 2017. *The effect of a performance-based financing program on HIV and maternal/child health services in Mozambique—an impact evaluation*. Health policy and planning, 32(10), pp.1386-1396. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5886140/	Payments to health facilities based on performance: Payment is based on the unit price of a service multiplied by the quantity of that service produced. An equity weight was applied to favour facilities in rural and hard-to-reach areas.	MNCH and HIV-PMTCT (prevention of mother to child) services in Mozambique	134 facilities	Activities and outputs	PMTCT, Paediatric HIV indicators vaccination coverages (e.g. number of women who got ARVs, received consults, babies who got the necessary tests, etc)	Majority of the 18 indicators responded to PBF, with at least half of indicators showing at least 50% improvement from baseline. However excluding pregnant women, the majority of adult HIV treatment indicators did not respond to PBF.

			Strong financial penalties are in place to dissuade inaccurate reporting. PBF earnings are allocated to facility investment (40%) and salary top-ups (60%). Salary top-ups are distributed among health facility staff based on a pre-determine					

			d criteria including years of experience and level of education. According to internal program data, salary top-ups account for between 20 and 50% of an average health workers salary, and facility investments account for approximately 50%					

			of total facility operating costs.					
Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. (2011) (Lancet) Available:	Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an	Basinga, P., Gertler, P.J., Binagwaho, A., Soucat, A.L., Sturdy, J. and Vermeersch, C.M., 2011. *Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an*	National supply side PBF program implemented at health facility level. The authors isolated	MNCH services in Rwanda as provided by facilities	166 facilities	Activities and outputs	Prenatal care visits, institutional deliveries, quality of prenatal care, child preventative visits, immunisation	23% increase in institutional deliveries in intervention group, 56% increase in preventative care visits for 0-23 months age group, 132% increase in preventative care

https://pubmed.ncbi.nlm.nih.gov/21515164/	impact evaluation	*impact evaluation.* The Lancet, 377(9775), pp.1421-1428. Available: https://pubmed.ncbi.nlm.nih.gov/21515164/	the incentive effect from the resource effect by increasing comparison facilities' input-based budgets by the average P4P payments made to the treatment facilities.					visits for 23-59 months age group, increased prenatal care quality measured by Rwandan prenatal clinical guidelines

Gertler P, Vermeersch C. (2013) (NBER Work Pap Ser.) Available: https://www.nber.org/system/files/working_papers/w19046/w19046.pdf	Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes	Gertler, P. and Vermeersch, C., 2013. *Using performance incentives to improve medical care productivity and health outcomes* (No. w19046). National Bureau of Economic Research. Available: https://www.nber.org/system/files/working_papers/w19046/w19046.pdf	National supply side PBF program implemented at health facility level. The authors used a control and comparison group, given the same budget, but with one side dependent on performance and the other not. The incentive structure	MNCH services in Rwanda as provided by facilities	10 districts in the treatment group and 9 in the comparison group	Activities, outputs and outcomes	Health worker productivity, child health outcomes	Substantial improvements in child health outcomes (weight and height), provider incentives led to 20% increase in productivity, evidence of complementarity between the incentive and knowledge (skill) of health care providers

			focuses not just on treating more patients, but also on providing more patients with higher quality of care (which has both a structural and process assessment aspect).					

Bonfrer I, Soeters R, Van de Poel E, Basenya O, Longin G, van de Looij F, et al (2013) (Health Affairs) Available: https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0081	Introduction of performance-based financing in Burundi was associated with improvements in care and quality	Bonfrer, I., Soeters, R., Van de Poel, E., Basenya, O., Longin, G., van de Looij, F. and van Doorslaer, E., 2013. *Introduction of performance-based financing in Burundi was associated with improvements in care and quality.* Health Affairs, 33(12), pp.2179-2187. Available: https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0081	Supply side PBF program. Facilities receive payments based on the quantity and quality of care. The government implemented a staggered rollout of this hence allowing for this study.	MNCH services in Burundi provided by facilities	2006: four provinces (two intervention and two control); from 2008: eleven provinces	Activities, outputs and outcomes	Institutional delivery, ANC services, vaccination coverage, ITN coverage, child illness episodes, waiting time. These are quantity indicators but facilities can also receive a quality bonus of up to 25%, as assessed by local regulatory authorities on a checklist (this has many items, focused on both structural and process quality)	PBF increased the probability of institutional deliveries by 21%, utilisation of antenatal care by 7% and the use of modern family planning by 5%

Binyaruka, P. and Borghi, J., 2017. Tropical Medicine and International Health, Volume 22 No. 1 pp. 92-102, January 2017. Available: https://onlinelibrary.wiley.com/doi/epdf/10.1111/tmi.12809	Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania	Binyaruka, P. and Borghi, J., 2017. *Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania*. Tropical Medicine & International Health, 22(1), pp.92-102. Available: https://onlinelibrary.wiley.com/doi/epdf/10.1111/tmi.12809	Incentive payout for the avoidance of stockouts result payouts to dispensaries, health centres and hospitals. Incentive payouts at the facility level include bonuses to staff (equivalent to 10% of their monhtly salary) and funds that can be used to	RMCH medical commodities in Tanzania as provided by local government facilities	75 intervention and 75 comparison facilities	Inputs	Availability of RMCH medicines, medical supplies and functioning equipment, and the stockouts of medicines and supplies at the facility.	P4P was associated with an 8.4% point increase in the availability of all 37 medicines combined. There were however varying results within this.

			procure drugs and supplies for facility improvement (10% of the total in hospitals and 25% in lower level facilities)					

de Walque D, Robyn PJ, Saidou H, Sorgho G, Steenland M. (2017) Available: https://openknowledge.worldbank.org/handle/10986/27969	Looking into the Performance-Based Financing Black Box: Evidence from an Impact Evaluation in the Health Sector in Cameroon	De Walque, D., Robyn, P.J., Saidou, H., Sorgho, G. and Steenland, M., 2017. *Looking into the performance-based financing black box: evidence from an impact evaluation in the health sector in Cameroon.* World Bank Policy Research Working Paper, (8162). Available: https://openknowledge.worldbank.org/handle/10986/27969	Payment of health facility bonus linked to volume and quality of services delivered in 14 districts. The authors tried to fill some gaps on how RBF works by isolating the role of explicit financial incentives as opposed to additional	MNCH services in Cameroon provided by facilities	4 health facilities	Activities and outputs	Child and maternal vaccinations, use of modern family planning, antenatal care visits, facility-based deliveries, patient satisfaction. An additional quality bonus provided an increase of up to 30% of the total payment based on health service quantity (standardised checklist, link provided, could not acess). Additionally, an equity bonus was included in the calculation of performance payments. The equity bonus was paid to health facilities that faced serious structural problems making	P4P efficient in bringing payments and funding to provider level, leading to increased coverage of MNCH and structural measures of quality of care; no difference in MNCH outcomes between health districts that were given P4P plus autonomy and those given incentive not attached to performance plus autonomy

			funding not linked to performance, as well as separating the impact of enhanced supervision and monitoring. District and regional managers also receive bonus payments per cycle based on the performance of facilities in their				service provision more challenging.	

			district or region.					

Shapira G, Kalisa I, Condo J, Humuza J, Mugeni C, Nkunda D, et al. (2017) Available: https://documents1.worldbank.org/curated/en/573571494939902839/pdf/WPS8059.pdf	Effects of Performance Incentives for Community Health Worker Cooperatives in Rwanda	Shapira, G., Kalisa, I., Condo, J., Humuza, J., Mugeni, C., Nkunda, D. and Walldorf, J., 2017. *Effects of performance incentives for community health worker cooperatives in Rwanda.* World Bank Policy Research Working Paper, (8059). Available: https://documents1.worldbank.org/curated/en/573571494939902839/pdf/WPS8059.pdf	Complementary community PBF program that rewarded community health worker cooperatives for the utilisation of five targeted MNCH services by their communities. When cooperatives received payments through the CPBF program, a	MNCH services in Rwanda as provided by health worker cooperatives	19 districts	Activities and outputs	Complete reporting, family planning consults, antenatal care, growth monitoring	The analysis finds no impact of the performance payments on coverage of the targeted services, attitudes and behaviors of community health workers, or outcomes at the cooperative level. No synergies are found between the scheme and a demandside, in-kind transfer intervention that was independently effective in increasing coverage rates of targeted services.

			maximum of 30 percent could be immediately distributed to cooperative members while the rest had to be invested in income-generating activities of the cooperatives' choosing. Revenues from these entrepreneurial activities,					

			in turn, could be reinvested or distributed as dividends to the members.					

Brenner, S., De Allegri, M., Kambala, C., Lohmann, J., Moszyk, D., Mazalale, J., & McMahon, S. (2016). http://sphfm.medcol.mw/wp-content/uploads/2016/07/Final-Results-Report-1.pdf	Final Results of the RBF4MNH Impact Evaluation	Brenner, S., De Allegri, M., Kambala, C., Lohmann, J., Moszyk, D., Mazalale, J. and McMahon, S., 2016. *Final results of the RBF4MNH impact evaluation*. Available: http://sphfm.medcol.mw/wp-content/uploads/2016/07/Final-Results-Report-1.pdf	Supply and demand-side incentives. Supply side incentives: Financial rewards to health facilities attaining pre-defined indicators of clinical and organizational performance in labour, delivery, and newborn care. The financial	Health facilities	18 facilities		The goal of RBF was to enable more women, particularly from poor rural areas, to deliver in health facilities, and for those facilities to offer better quality maternal and neonatal care services.	While we observed no statistically meaningful difference in overall clinical quality, improvements were detected on single quality indicators, but not necessarily on the overall process within which this indicator was placed. This situation is likely linked to the general difficulties RBF programs face in setting and verifying adequate quality indicators and targets for clinical performance. In other words, indicators that are

			allocation was to be utilised depending on the hospital's needs for investment in promoting quality maternal and newborn health care. Demand side: cash transfer to pregnant women.					readily measured and verified are possibly not the same factors that create differences in the process of delivery care. In addition, optimal clinical performance is dependent on a variety of optimal service inputs. Both quantitative and qualitative findings point towards the fact that the most crucial input factor – the number of qualified health workers – remain still short at many health centers.

Antony, M., Bertone, M. P., & Barthes, O. (2017).	Exploring implementation practices in results-based financing: the case of the verification in Benin	Antony, M., Bertone, M.P. and Barthes, O., 2017. *Exploring implementation practices in results-based financing: the case of the verification in Benin*. BMC health services research, 17(1), pp.1-12.	Supply side RBF program focusing on the verification of results. Facilities received payment per quantity / quality indicator based on 100 indicators. These indicators focused on the quality of the facilities' equipment and service	Health facilities in Benin	8 districts	analysis focuses on the actual practices of quantitative, qualitative and community verification.	Health facilities' productivity and quality of healthcare	

			delivery environment.					
Celhay PA, Gertler PJ, Giovagnoli P, Vermeersch CMJ. (2015)	Argentina : Can Short Term Incentives Change Long Term Behavior?	Celhay, P.A., Gertler, P.J., Giovagnoli, P. and Vermeersch, C.M., 2015. *Argentina-Can short term incentives change long term behavior?* (No. 100060, pp. 1-4). The World Bank.	provincial payments partly determined by performance	province	37 facilities	outcome level	Outcome indicators	Improved child outcomes
Chung, S., Palaniappan, L., Wong, E., Rubin, H., Luft, H. (2009)	Does the frequency of pay-for-performance payment matter? Experience from a randomized trial	Chung, S., Palaniappan, L., Wong, E., Rubin, H. and Luft, H., 2010. *Does the Frequency of Pay-for-Performance Payment Matter?—Experience from a Randomized Trial.* Health Services	PFP program paying the providers of primary care a maximum of U.S.$5,000/year bonus	Providers of primary care	179 providers		Quality measure scores	There was no difference in the average quality measure scores between the two groups receiving the bonus payments at different intervals in the year.

		Research, 45(2), pp.553-564.	(either paid quarterly or annually)					
Custers, T., Hurley, J., Klazinga, N. S., & Brown, A. D. (2008)	Selecting effective incentive structures in health care: A decision framework to support health care purchasers in finding the right incentives to drive performance	Custers, T., Hurley, J., Klazinga, N.S. and Brown, A.D., 2008. *Selecting effective incentive structures in health care: A decision framework to support health care purchasers in finding the right incentives to drive performance*. BMC health services research, 8(1), pp.1-14.	Literature review on RBF to compose a framework for making decisions on choosing and designing the RBF incentive to utilise.	Providers of health services	85 articles		The review looked at various measures of quality when reviewing the literature	A number of contextual influences need to be considered when designing successful RBF schemes. Most incentives showed little impact on performance improvements.

Eijkenaar, F. (2013).	Key issues in the design of pay for performance programs	Eijkenaar, F., 2013. *Key issues in the design of pay for performance programs*. The European Journal of Health Economics, 14(1), pp.117-131.	Literature review discussing the main design issues in PFP programs					Design measures impact quality outcomes
Gergen, J., Josephson, E., Vernon, C., Ski, S., Riese, S., Bauhoff, S., & Madhavan, S. (2018)	Measuring and paying for quality of care in performance-based financing: experience from seven low and middle-income countries (Democratic Republic of Congo, Kyrgyzstan, Malawi,	Gergen, J., Josephson, E., Vernon, C., Ski, S., Riese, S., Bauhoff, S. and Madhavan, S., 2018. *Measuring and paying for quality of care in performance-based financing: experience from seven low and middle-income countries*	Literature review discussing the challenges and lessons to be learned from PBF schemes in various LMICs	Providers of health services in various LMICs	8 case studies		Various structural and process measures of quality	There was some improvement in performance after the implementation of the PBF programs in the various LMICs.

	Mozambique, Nigeria, Senegal and Zambia).	*(Democratic Republic of Congo, Kyrgyzstan, Malawi, Mozambique, Nigeria, Senegal and Zambia)*. Journal of global health, 8(2).						
Gergen, J., Rajkotia, Y., & Ravishankar, N. (2018).	The good, the bad, and the disruptive of performance-based financing on the Mozambican health system: results from a process evaluation	Gergen, J., Rajkotia, Y. and Ravishankar, N., 2018. *The good, the bad, and the disruptive of performance-based financing on the Mozambican health system: results from a process evaluation*. Sociedade e Cultura, 21(2).	A process evaluation of a PBF program focused on providing bonuses to health facility workers. The study utilised qualitative methods (interviews and group discussion	Health workers and facility administrators in health facilities within Mozambique	24 facilities		Perceived impacts of the PBF program on four main areas, namely service delivery, health financing, human resources and governance of the health facilities.	PBF improved the facilities' work environment through improved local financial capacity and autonomy, resulting in greater planning

			s) to collect data in 24 PBF health facilities from 60+ health workers and facility administrators.					
KIT, J. T., KIT, A. C., KIT, P. V., & WHO, R.	LEARNING LESSONS ON IMPLEMENTING PERFORMANCE BASED FINANCING, FROM A MULTI-COUNTRY EVALUATION KIT	KIT, J.T., KIT, A.C., KIT, P.V. and WHO, R.E., *LEARNING LESSONS ON IMPLEMENTING PERFORMANCE BASED FINANCING, FROM A MULTI-COUNTRY EVALUATION KIT* (ROYAL TROPICAL INSTITUTE) In collaboration with Cordaid and WHO.	A synthesis report that explores the lessons learned on design, implementation and effects of financial incentives in the	Providers of health services in countries in Sub-Saharan Africa	5 countries		Various indicators of quality were reviewed, including perceived quality of care by patients and provision of equipment, medicines and infection control	PBF schemes may be better than traditional funding methods at improving the productivity and quality of healthcare service delivery

			form of performance based financing in the health sector, as supported in Sub–Saharan Africa by the two Dutch NGO's Cordaid and HealthNet TPO					
Lorincz, I. S., Lawson, B. C., & Long, J. A. (2013).	Provider and patient directed financial incentives to improve care and outcomes for patients with diabetes	Lorincz, I.S., Lawson, B.C. and Long, J.A., 2013. *Provider and patient directed financial incentives to improve care and outcomes for patients with diabetes*. Current	Literature review focusing on the impact that demand-side PBF programs				Tests of patient health such cholesterol tests, blood pressure tests and weight loss were reviewed, as well as process quality tests.	Demand-side PBF schemes improved positive patient behaviour changes, however the long term effects of the programs are uncertain.

		diabetes reports, 13(2), pp.188-195.	have had on patients with diabetes					
Cole, M. S., Boydell, V., Bellows, B., & Hardee, K. (2018)	Mapping the Extent to Which Performance-Based Financing (PBF) Programs Reflect Quality, Informed Choice, and Voluntarism and Implications for Family Planning Services: A Review of PBF Operational Manuals.	Cole, M.S., Boydell, V., Bellows, B. and Hardee, K., 2018. *Mapping the Extent to Which Performance-Based Financing (PBF) Programs Reflect Quality, Informed Choice, and Voluntarism and Implications for Family Planning Services: A Review of PBF Operational Manuals*.	Literature review focusing on the extent to which family planning principles are currently considered within PBF programs		23 programmes		Existing family planning indicators	Many RBF policies do not adequately address rights-based patient-centred quality

Maddox, K. E. J., Sen, A. P., Samson, L. W., Zuckerman, R. B., DeLew, N., & Epstein, A. M. (2017).	Elements of program design in medicare's value-based and alternative payment models: a narrative review	Maddox, K.E.J., Sen, A.P., Samson, L.W., Zuckerman, R.B., DeLew, N. and Epstein, A.M., 2017. *Elements of program design in medicare's value-based and alternative payment models: a narrative review*. Journal of general internal medicine, 32(11), pp.1249-1254.	Review of medicare payment models and literature, attempting to decide which payment model is superior		7 models		Efficiency of each model in reaching specific performance targets	Each model has its own pro's and con's. There is no single model that is significantly better than another
Mantzari, E., Vogt, F., Shemilt, I., Wei, Y., Higgins, J. P., & Marteau, T. M. (2015).	Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis.	Mantzari, E., Vogt, F., Shemilt, I., Wei, Y., Higgins, J.P. and Marteau, T.M., 2015. *Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis*. Preventive	Systematic review and meta-analysis	Adult individuals	34 articles		Various health indicators including smoking status, alcohol consumption and physical activity.	Personal financial incentives can help induce positive behaviour changes in individuals, and help lower healthcare inequalities.

		medicine, 75, pp.75-85.						
Rosenthal, M.B., Dudley, R.A (2007)	Pay-for-performance: will the latest payment trend improve care?	Rosenthal, M.B. and Dudley, R.A., 2007. *Pay-for-performance: will the latest payment trend improve care?*. Jama, 297(7), pp.740-744.	Commentary on the care improvements initialised by RBF schemes	Providers			Clincal appropriateness of care	Economic theories should be considered in the design process to ensure that quality is improved
Seddon, M.E., Marshall, M.N., Campbell, S.M., Roland, M.O (2001)	Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand.	Seddon, M.E., Marshall, M.N., Campbell, S.M. and Roland, M.O., 2001. *Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand.* BMJ Quality &	Systematic review and meta-analysis	Health care facilities in the UK, Australia and New Zealand	90 articles		Various measures such as access to healthcare, clinical effectiveness, interpersonal effectiveness, equity and efficiency	

		Safety, 10(3), pp.152-158.						
Singh, N. S., Kovacs, R. J., Cassidy, R., Kristensen, S. R., Borghi, J., & Brown, G. W. (2020).	A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries	Singh, N.S., Kovacs, R.J., Cassidy, R., Kristensen, S.R., Borghi, J. and Brown, G.W., 2020. *A realist review to assess for whom, under what conditions and how pay for performance programmes work in low-and middle-income countries.* Social Science & Medicine, p.113624.	Systematic literature review focusing on the effectiveness of PFP schemes in LMICs	Health care facilities in LMICs	112 articles		Various measures including availability of equipment and medicines and effectiveness of patient-provider interactions	The study found that PFP programs improved health facilities' access to resources, interactions with patients, adherance to healthcare guidelines and resulted in greater patient satisfaction. Contextual factors also influenced the effectiveness of PFP in enhancing the healthcare system.

Patel, S. (2018)	Structural, institutional and organizational factors associated with successful pay for performance programmes in improving quality of maternal and child health care in low and middle income countries: a systematic literature review	Patel, S., 2018. *Structural, institutional and organizational factors associated with successful pay for performance programmes in improving quality of maternal and child health care in low and middle income countries: A systematic literature review*. Journal of global health, 8(2).	Systematic literature review	Health care facilities in LMICs	13 articles		Multiple structural and process quality measures, including availability of healthcare equipment, medicines, adherence to medical guidelines, staff skill levels	Some measures of maternal and child healthcare quality improved, and various statistics such as the number of prenatal visits, institutional deliveries and preventative visits for young children increased.

Education Literature

Name of paper	Author	Type of incentive/RBF	Target group	Sample size	Level in results chain where disbursement is triggered in health (compare to educ) (inputs, activities, outputs, outcomes, impact)	Measures of quality	Impact on quality	
An Analysis of Foundati	Muhammad Nadeem Anwar, Asma Khizar & Raheel Haq (2018). Bulletin of Education	PEF provides monthly subsidies to support low-		3500 schools in 36 districts	Outcomes	Infrastructure and facilities, pupil numbers, outcomes	Improved enrolment, grade transition and	

on Assisted Schools Program of Punjab as a Mechanism Influencing Pupil Cohort	and Research August 2018, Vol. 40, No. 2 pp. 1-12. https://files.eric.ed.gov/fulltext/EJ1209787.pdf	cost private schools in the region, conditional upon them meeting specified target pass rates on standardized exams. PPP		across Punjab, Pakistan			survival rates of learners	
Teacher pay for performance: Does it really work?	Beteille, T., Breeding, M. & Evans, D. (2021). https://blogs.worldbank.org/education/teacher-pay-performance-does-it-really-work	Out of 15 evaluated programs in low and middle-income countries, 9 provided individual monetary incentives, 5 provide	Teacher PFP programs in low and middle-income countries	Meta-study of 15 evaluations		Learning outcomes	Effects on student learning outcomes vary substantially. Across programs (72 reported outcomes of impacts on student test scores in 15 evaluations), effect sizes range in magnitude with a minimum effect of -	

		group-based monetary incentives, and 1 provided in-kind group-based incentives. One program, provided both individual and groupbased monetary incentives.					0.08 Standard Deviations (SD) and a maximum effect of 0.32 SD increase in student test scores. The median reported effect size is a 0.06 SD increase. Most programs produce a mix of outcomes across different measures of test scores and subjects, most of which are substantively and statistically insignificant when it comes to increasing student test scores	
Evaluation of the	Cambridge Education. (2015).	To encourage increases in	Grade 10	EGSECE sitters:			The evaluation was unable to detect	

pilot project of results-based aid in the education sector in Ethiopia	https://reliefweb.int/sites/reliefweb.int/files/resources/Eval-pilot-project-result-based-aid-Education-Ethiopia.pdf	number of sitters and passers, DFID offered the Government of Ethiopia (GoE) up to £10 million per year for each of three years for increases in the number of students sitting for and passing the EGSECE. The amounts to be provided per additional sitter and passer are shown in Table E.1, with higher	students (boys and girls and especially among students in Ethiopia's developing regional states (DRS), commonly referred to as emerging regions) sitting	1 639 935 EGSECE passers: 1 385 713			evidence that the increases in the numbers of the grade 10 examination sitters and passers. This is because of the absence of a real counterfactual.	

		amounts for girls than for boys and for students in the emerging regions. Reward payments would be based on the numbers of additional sitters and passers within each region compared with the number of sitters and passers in each region the previous year, thus using an 'adjusting' or	for the Ethiopian General Secondary Education Certificate Examination (EGSECE) - 10th Grade National Examination					

		rolling baseline1 . The reward payments for additional sitters would be provided irrespective of their performance on the EGSECE.						
Paying for Education Outcomes at Scale in India	Gustafsson-Wright, E., & Boggild-Jones, I. (2019). Brookings Institution (USA). Center for Universal Education. https://files.eric.ed.gov/fulltext/ED602924.pdf	Educate Girls DIB		Students in Grades 3-5 in 166 treatment schools (7,318 students were in Grades 1-5 in treatment schools		Learning gains of students enrolled in grades 3-5 and enrollment of out-of-school girls	The project had enrolled 768 of the out-of-school girls identified at baseline, or 92%, against a target of 79%. The intervention achieved student learning outcome gains equivalent to 160% of the target.	

				at baseline) in Rajasthan's Bhilwara district			On average, students in EG schools gained an additional 1.08 ASER levels compared to students in control schools ($p < 0.01$). Differences in aggregate learning gains between treatment and control schools were much greater in Year 3 (+6,045 learning levels) than in Year 2 (+1,434 levels) or in Year 1 (+1,461 levels) While program impact was significant in all subjects, it was ~3 times larger in	

							Math and English than in Hindi. Students in Educate Girls' program gained an additional 0.31 standard deviations in test scores over the course of the three-year evaluation.	
Quality Education India Development Impact Bond: A case study produced as part of the independent	Erskine, C. (2019). Ecorys. https://s3.eu-west-2.amazonaws.com/golab.prod/documents/Quality_Education_India_DIB_Case_Study_Report.pdf	Quality Education India DIB	Government and low-fee private schools in the states of Delhi, Gujarat, Maharashtra,	200,000 students in Grades 1-8		Learning outcomes in literacy and numeracy.	After the first year of implementation, 40% of schools in the DIB met or exceeded their learning targets. Four service providers reached 104,833 students across 711 schools over the two-year period. On average, the cumulative achievement across	

evaluation of the Department for International Development's Development Impact Bond Pilot Programme			and Uttar Pradesh				all grades over two years demonstrates that QEI DIB schools substantially overperformed relative to targets.	
Teacher Performance Pay: Experimental Evidence from India	Muralidharan, K., and Sundararaman, V. (2011). Journal of Political Economy, University of Chicago Press, vol. 119(1), pages 39 - 77	Teacher incentive program two types of teacher performance pay (group bonuses based on school	Government-run schools in rural Andhra Pradesh, India	300 government-run schools in rural AP with 100 schools each in the			At the end of two years of the program, students in incentive schools performed significantly better than those in control schools by 0.28 and 0.16 standard deviations	

		performance, and individual bonuses based on teacher performance). The program provided bonus payments to teachers based on the average improvement of their students' test scores in independently administered learning assessments (with a mean bonus of 3%		group and individual incentive treatment groups and 100 schools serving as the comparison group			in math and language tests respectively. They scored significantly higher on "conceptual" as well as "mechanical" components of the tests, suggesting that the gains in test scores represented an actual increase in learning outcomes. Incentive schools also performed better on subjects for which there were no incentives, suggesting positive spillovers. Group and individual incentive schools performed equally well in the first year	

		of annual pay)					of the program, but the individual incentive schools outperformed in the second year. Incentive schools performed significantly better than other randomly-chosen schools that received additional schooling inputs of a similar value.	
Results-Based Financing in Education: Learning from What Works	Lee, J., & Medina, O. (2019). https://documents1.worldbank.org/curated/en/915061548222619389/pdf/Results-Based-Financing-in-Education-Learning-from-What-Works.pdf	Teacher incentive for attendance; learner incentives (incentivising learner outputs and outcomes);	Experimental or quasi-experimental impact evaluations in developing	42 impact evaluation papers; 8 reviews of evidence and meta-analyses; 46 staff		Teacher attendance; learning outcomes; learner attendance; learner enrolment	Teacher incentives can but do not always improve teacher attendance and student learning. The design of the incentive scheme and the context matter. The effects are larger and more	

		school-based grants	countries; meta-analyses; staff from development agencies;	from development agencies who design, implement, and evaluate RBF interventions and programs in the education sector (surveys conducted with all 46 and in-depth interviews carried out with 19 of			positive in developing country contexts. Student and family incentives (such as CCTs, for instance) has a good track record of reducing school dropout and increasing school attendance, though the evidence for its effects on student learning is more mixed. The evidence base on the effectiveness of performance based grants is still quite limited. There is growing evidence that	

				these staff)			combining different RBF interventions within the same program can generate results that go beyond the sum of any two interventions alone. According to the available evidence, the overall effect of grants on learning outcomes is mixed.	
The impact of education programmes on learning and school participa	Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., Maxwell Stamp, Jobse, H., Schmidt, T., & Jimenez, E. (2016). International Initiative for Impact Evaluation (3ie).		Impact and process evaluations; qualitative research studies	238 impact evaluations which examined the effects of 216 programmes		Primary: Enrolment; attendance; completion; learning outcomes Secondary: teacher attendance; teacher performance	Cash transfers had the largest and most consistent positive effects on increasing school enrolment, reducing dropouts and improving completion. However, on	

tion in low- and middle-income countries (2016th ed.)	https://doi.org/10.23846/SRS007		and education programmes in LMICs	across 52 LMICs (the number of impact evaluations is higher than the number of programmes as multiple impact evaluations were conducted for some programmes. 121 qualitative			average, they have not improved learning outcomes. Structured pedagogy programmes had the largest and most consistent positive effects on improving learning outcomes. These are programmes that typically provide customised curricula, new instructional approaches and teachers' training, and educational materials for students.	

				research studies and process evaluations				
Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990–2010)	Mizala, A., & Schneider, B.R. (2014). Governance: An International Journal of Policy, Administration, and Institutions, Vol. 27, No. 1, (pp. 87–109).	Collective and individual pay incentives for teachers (salary incentives)					Ongoing negotiation with the teacher union resulted in an institutionalized structure of incentive pay for teachers as well as widespread attitudes of sustained support among teachers for performance pay. The amounts of incentive pay by the late 2000s ranged roughly 15–25% of base pay, and the multiple incentives	402 685

							and associated evaluations shifted overall career expectations. Teachers entering the profession in the late 2000s had quite different career expectations from those who started earlier. The later entrants knew that they would be evaluated on a regular basis and that the evaluations could increase their earnings, or in the case of the worst performers in public schools, lead to sanctions and dismissal.	

							Moreover, some positive impacts of incentive schemes on educational outcomes can only emerge over the longer term as they shift the profile and expectations of teachers entering the profession.	
Improvements in the Quality of Basic Education: Chile's Experience.	Wales, J., Ali, A., Nicolai, S., Morales, F., and Contreras, D. (2014). Case Study Report: Education, Overseas Development Institute, London.	Performance-related incentives for teachers SNED (1995) introduced collective incentives for schools - teachers from winning schools awarded a				Teacher performance; learning outcomes	SNED: impact evaluations have found the results to be positive and analysis of secondary sources suggests that this may have been one of the most cost-effective initiatives of this period (1995). In the National	

		bonus equivalent to an additional month's salary for each teacher for a two-year period. National Teacher Evaluation Programme. Multi-faceted assessment system established for all teachers in municipal schools. Teachers to be evaluated at least once every four years. High					Teacher Evaluation Programme, teachers are assessed every four years. Those performing below standard are reassessed the following year, while those who achieve the basic level are assessed again after two years. All teachers falling into these categories are offered pedagogical training, for which municipalities are given the necessary funding. Teachers who are rated 'competent' or 'outstanding'	

		performing teachers become eligible for a series of performance -related incentives and peer-training schemes. Low performing teachers must retrain and can be made redundant for repeatedly poor performance. Pedagogical Excellence Allowance					are able to take further tests in their pedagogical and subject area, which will, if they perform well, enable them to receive the Variable Incentive for Individual Performance – representing an increase in salary of CLP 46,000 per month ($83.3016) that lasts between two and four years, depending on the teacher's test results. Chile, along with two OECD members, were the only countries to	

		and incentives for peer training.					improve by more than 20 points in reading assessments over 2000-2009; at the same time Chile demonstrated science improvements over 2006-2009 that were above the OECD average level. The country also witnessed a steady improvement in literacy rates over the last three decades. By 2009 literacy was almost universal amongst youth and adult populations and the country's primary level pupil to teacher ratio fell	

							impressively over 1999-2010 from 32 to 23.5.	
World Development Report 2018 (WDR 2018)—LEARNING to Realize Education's Promise	World Bank (2018). https://www.worldbank.org/en/publication/wdr2018	Financial incentives to teachers				Teacher motivation; learning outcomes	In India, students performed better in primary schools that provided teachers with financial incentives for higher reading and mathematics scores. Students also scored higher in science and social studies, despite no financial incentives being offered in those areas. In the United States, by contrast, teacher financial incentives did not improve test scores	

							in several states. However, large financial incentives for teachers did increase student learning in the District of Columbia, United States.46 In Mexico and Tanzania, teacher financial incentives were effective only in conjunction with another intervention. One interpretation of this scattered evidence is that financial incentives are most likely to be effective when teachers can take straightforward	

							actions to improve learning.	
From proof of concept to scalable policies: challenges and solutions, with an application.	Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. Journal of Economic Perspectives 31 (4): 73–102. https://doi.org/10.1257/jep.31.4.73		N/A	Pupils in early grades	This paper is an extensive reflection on taking TaRL from proof of concept to take-up by government in India, over a period of 10 years.	Outcomes	Test scores in reading and maths	Consistently found to improve test scores
Scaling up successfully: Lessons from Kenya's Tusome national	Piper, B., Destefano, J., Kinyanjui, E. & Ong'ele, S. (2018). Journal of Educational Change,19:293–321. https://doi.org/10.1007/s10833-018-9325-4		N/A	All early grade learners in Kenya	Nationally representative	N/A	Reading outcomes	Improved reading scores on a number of measures

literacy program.								
Independent evaluation of the UK Department for International Development's development impact bonds (DIBs) pilot programme – Full report	Cox, K., Ronicle, J., Lau, K., & Rizzo, S. (2019).	Development impact bond	Primary school learners in Delhi and Gujarat	Approximately 200 00 primary school children aged 5-11 across 600 schools in Delhi and Gujarat, India		Learning outcomes in literacy and numeracy		

World's largest education DIB reports first-year results.	Edwards, S. (2019). https://www.devex.com/news/world-s-largest-education-dib-reports-first-year-results-95216	DIB	Primary school learners in Delhi and Gujarat	Approximately 200 00 primary school children aged 5-11 across 600 schools in Delhi and Gujarat, India		Learning outcomes in literacy and numeracy	After the first year of the QEI DIB, 40% of participating schools either met or exceeded their targets for literacy and numeracy skills, however 60% of schools failed to outperform non-participating schools. A teacher training programme that was part of the QEI DIB underperformed in north Delhi and it was decided to discontinue this component of the DIB.	
Educate Girls	Kitzmüller, L., McManus, J., Shah,	DIB	Primary	Boys and girls in		Learning gains in primary school	By the end of the three-year	

Development Impact Bond: Final Evaluation Report.	N.B., & Sturla, K. (2018). IDinsight. https://s3.eu-west-2.amazonaws.com/golab.prod/documents/ID_Insight_2018_Educate_Girls_Development_Impact_Bond_-_Final_Evaluation_Report.pdf		school learners and out of school girls in rural Rajasthan	grades 3-5 (7318 students) and 837 out of school girls		learners and enrolment for out of school girls	programme, students had gained an additional 8,940 learning levels, representing 160% of the final target of 5,592. 768 out-of-school girls, representing 92% of the 837 eligible out-of-school girls were enrolled at the end of the programme	
Social impact bonds fund welfare projects: how South	Khan, Z. (2021). The Conversation. https://theconversation.com/social-impact-bonds-fund-welfare-projects-how-south-	SIB	2000 preschool-aged children in Delft and	Impact Bond Innovation Fund (IBIF) implemented in		Recruitment and retention of children; attendance of sessions; early learning	The project significantly over-achieved on the first target. However, though improvements were achieved, the early learning	

Africa's first two have done	africas-first-two-have-done-160883		Atlantis, Western Cape, South Africa	South Africa			outcome measure targets were missed.