ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Holzapfel, Sarah; Janus, Heiner

Working Paper

Improving education outcomes by linking payments to results: an assessment of disbursement-linked indicators in five results-based approaches

Discussion Paper, No. 2/2015

Provided in Cooperation with:
German Institute of Development and Sustainability (IDOS), Bonn

Suggested Citation: Holzapfel, Sarah; Janus, Heiner (2015) : Improving education outcomes by linking payments to results: an assessment of disbursement-linked indicators in five results-based approaches, Discussion Paper, No. 2/2015, ISBN 978-3-88985-665-4, Deutsches Institut für Entwicklungspolitik (DIE), Bonn

This Version is available at:
https://hdl.handle.net/10419/199452

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

d·i·e

Deutsches Institut für Entwicklungspolitik

German Development Institute

Discussion Paper 2/2015

Improving Education Outcomes by Linking Payments to Results

An Assessment of Disbursement-linked Indicators in five Results-based Approaches

Sarah Holzapfel
Heiner Janus

Improving education outcomes by linking payments to results

An assessment of disbursement-linked indicators in five results-based approaches

Sarah Holzapfel

Heiner Janus

Bonn 2015

Discussion Paper / Deutsches Institut für Entwicklungspolitik
ISSN 1860-0441

Die deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available at http://dnb.d-nb.de.

ISBN 978-3-88985-665-4

Sarah Holzapfel, Department "Bi- and Multilateral Development Cooperation", German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE)

E-mail: sarah.holzapfel@die-gdi.de

Heiner Janus, Department "Bi- and Multilateral Development Cooperation", German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE)

E-mail: heiner.janus@die-gdi.de

© Deutsches Institut für Entwicklungspolitik gGmbH
Tulpenfeld 6, 53113 Bonn
☎ +49 (0)228 94927-0
🖷 +49 (0)228 94927-130
E-Mail: die@die-gdi.de
www.die-gdi.de

Abstract

In results-based approaches, funding is linked to pre-agreed results that are defined in the form of indicators. Disbursements only take place once progress in the indicators has been verified. This places high requirements on the quality of indicators used. Different development actors have started implementing results-based approaches, yet little attention has been paid to potential advantages and disadvantages of the specific indicators that are used. The paper addresses this gap by first conceptualising a typology of indicators and devising criteria for assessing the quality of indicators. The typology and criteria are then applied to five results-based pilot programmes in the education sector in developing countries (Ethiopia, Rwanda, Sri Lanka, Tanzania). A comparison of the indicators used across these programmes provides insights into how indicators for results-based approaches can be selected in a more informed manner in the future.

Key words: results-based approaches, results-based aid, development cooperation, disbursement-linked indicators, education

Acknowledgements

We would like to thank Emma Fan, Aimée Hampel-Milagrosa, Niels Keijzer, Stephan Klingebiel, Constantine Manda, Donald Menzies and Youdi Schipper for extremely valuable comments and suggestions. We would also like to thank all interviewees who contributed to this study. The views and opinions expressed in this paper, as well as any errors and omissions, are our own.

Bonn, February 2015

Sarah Holzapfel

Heiner Janus

Contents

Abbreviations

1 Introduction 1

2 Classification of indicators in results-based approaches 2

2.1 Indicators at different levels of the results chain 4

2.2 Direct and indirect indicators 5

2.3 Quantitative and qualitative indicators 6

3 Criteria for assessing indicators used to disburse payments 7

3.1 Focus on results 7

3.2 Control 8

3.3 Financial incentives 9

3.4 Measurability and verifiability 11

3.5 Unintended consequences 12

4 Overview of results-based approaches in education 14

4.1 Analysis of the indicators used in five selected programmes 16

4.1.1 World Bank “Big Results Now in Education Program” Tanzania 16

4.1.2 ADB “Education Sector Development Programme” Sri Lanka 22

4.1.3 DFID “Results Based Aid Pilot – Education Sector Programme” Rwanda 27

4.1.4 DFID “Pilot Project of Results Based Aid in the Education Sector” Ethiopia 32

4.1.5 Twaweza “KiuFunza Local Cash on Delivery” Tanzania 37

4.2 Comparison of indicators used in the five selected programmes 40

4.2.1 Focus on results 41

4.2.2 Control 42

4.2.3 Financial incentives 43

4.2.4 Measurability and verifiability 44

4.2.5 Unintended consequences 45

5 Conclusions and recommendations 46

Bibliography 49

Figures

Figure 1:	The results chain	4
Figure 2:	Outcome orientation, numbers of indicators and programme budgets	41

Tables

Table 1:	Typology of indicators	3
Table 2:	Criteria to assess the quality of disbursement-linked indicators	7
Table 3:	Results-based approaches in education	16
Table 4:	World Bank “Big Results Now in Education Program” Tanzania	17
Table 5:	The six intermediate objectives/clusters of DLIs	18
Table 6:	ADB “Education Sector Development Program” Sri Lanka	22
Table 7:	Intermediate objectives of the ADB Education Sector Development Programme Sri Lanka	23
Table 8:	DFID “Results Based Aid Pilot – Education Sector Programme” Rwanda	28
Table 9:	DFID “Pilot Project of Results Based Aid (RBA) in the Education Sector” Ethiopia	33
Table 10:	Payment structure	35
Table 11:	Twaweza “KiuFunza Local Cash on Delivery” Tanzania	38

Annex Tables

Annex Table 1:	World Bank “Big Results Now in Education Program” Tanzania	55
Annex Table 2:	ADB “Education Sector Development Program” Sri Lanka	59
Annex Table 3:	DFID “Results Based Aid Pilot – Education Sector Programme” Rwanda	62
Annex Table 4:	DFID “Pilot Project of Results Based Aid in the Education Sector” Ethiopia	62
Annex Table 5:	Twaweza “KiuFunza Local Cash on Delivery” Tanzania	62

Abbreviations

3R	Reading, writing and arithmetic
ADB	Asian Development Bank
BRN	Big Results Now (Tanzania)
BRNEd	Big Results Now in Education Program (Tanzania)
CEFR	Common European Framework of Reference for Languages
CELD	Center for Education Leadership Development
CGD	Center for Global Development
COD Aid	Cash on Delivery Aid
DAC	Development Assistance Committee
DFID	Department for International Development (United Kingdom)
DIE	Deutsches Institut für Entwicklungspolitik / German Development Institute
DLI	Disbursement-linked indicator
DTP3	Diphtheria-tetanus-pertussis
ESDP	Education Sector Development Programme (Sri Lanka)
GAVI	Global Alliance for Vaccines and Immunisation
GBP	British pound
GCE	General Certificate of Education
GFATM	Global Fund to Fight AIDS, Tuberculosis and Malaria
GoR	Government of Rwanda (GoR)
IMF	International Monetary Fund
LGA	Local Government Authority
M&E	Monitoring and evaluation
MOE	Ministry of Education
NGO	Non-governmental organisation
OECD	Organisation for Economic Co-operation and Development
PforR	Program for Results
RBA	Results-based aid
RBF	Results-based finance
RBL	Results-based lending
RCT	Randomised control trial
SMTSU	Sector Monitoring and Technical Support Unit
TZS	Tanzanian shilling
UK	United Kingdom
UNDP	United Nations Development Programme
USD	United States dollar
WB	World Bank
wpm	Words per minute

1 Introduction

Improving development cooperation towards achieving results in the form of observable improvements in peoples' lives remains high on the international agenda. As part of a larger discussion on how to deliver and demonstrate development results, donors and partner countries have begun to implement results-based approaches.[1] A common element of these results-based approaches is to link the level of funding directly to pre-agreed results that are measured through indicators. Based on a regular verification of these indicators, ex-post payments are then made. This innovative approach promises several potential advantages over traditional aid modalities, such as improved incentives, accountability, monitoring and recipient discretion. But whether results-based approaches deliver higher returns than traditional aid modalities has yet to be confirmed in practice since most pilot programmes are still at an early stage of implementation.

A central question for results-based approaches is: What are good indicators and how can they be selected? The choice and definition of disbursement-linked indicators is critical because financial incentives can be extremely powerful. If indicators are poorly defined or incomplete, the results may not be fully measurable, which would make it difficult to pay for performance. In addition, if the intended results are not clearly defined ex-ante or seem unattainable, the recipient may not have a sufficient incentive to attempt to achieve the results. Only if indicators are adequately selected and create the right incentives can a results-based approach be successful and contribute to the achievement of long-term development progress.

The existing results-based pilot programmes are being implemented by various actors, including the World Bank (WB), the Asian Development Bank (ADB), the United Kingdom Department for International Development (DFID) and non-governmental organisations (NGOs). Each actor has introduced a different version of results-based approaches, and thus different ways of using indicators. Some approaches, such as the World Bank's Program-for-Results (PforR) or the ADB's Results-Based Lending (RBL) instrument use multiple indicators at different levels of the results chain (i.e., input-, activity-, output- or outcome-level indicators). Other approaches focus on a specific level of the results chain and use a limited number of indicators. For example, the Cash on Delivery Aid (COD Aid) approach developed by the Center for Global Development and piloted by DFID, uses few indicators and relies on outcome-level indicators only.

Although the choice of adequate indicators is critical for the success of results-based approaches, only few studies have addressed this topic. Existing literature focuses on developing a typology of indicators or defining quality criteria for specific types of results-based approaches (see for example Asian Development Bank [ADB], 2013b, p. 20; Gelb & Hashmi, 2014; Savedoff & Martel, 2011; World Bank, 2012, pp. 34-35). While Gelb and Hashmi (2014) provide a first overview of types of indicators used across results-based operations of the World Bank, there is a lack of studies that empirically analyse the quality of indicators used in ongoing programmes.

1 Result-based approaches generally fall into one of two categories: result-based aid (RBA) and results-based finance (RBF) (Janus, 2014; Klingebiel, 2012; Klingebiel & Janus, 2014; Pearson, Johnson, & Ellison, 2010). RBA takes the form of a partnership between a donor and a government, whereas RBF uses domestic funding in a contractual arrangement between a government as the funder and an implementing actor (sub-national government, non-governmental organisation or private sector).

This paper addresses the identified research gap by developing a typology of indicators and a list of criteria to assess the quality of indicators used in different types of results-based approaches. Based on the typology and the criteria, the paper analyses indicators used in five selected programmes in the education sector. Two of the selected programmes are implemented by development banks, a World Bank pilot programme and an ADB pilot programme. Three programmes are based on the Cash on Delivery Aid model (Birdsall & Savedoff, 2010) and are implemented by DFID and a local non-governmental organisation (NGO) in Tanzania called Twaweza.

The analysis provides specific insights into which criteria indicators have to fulfil when aid disbursements are based on indicators in results-based approaches. On a more general level, the analysis will also better inform the selection of indicators for the design of upcoming results-based approaches. The quality of indicators used critically determines the effectiveness of results-based approaches and might even determine the overall value of results-based approaches as an aid modality. The education sector is well suited for the analysis because it is one of the few sectors where several results-based approaches are already in operation and where first implementation experiences can be observed.

For the purpose of the analysis, we draw on existing literature on indicators and results-based approaches as well as on guidance documents by various different development agencies. To gain additional information on the five programmes reviewed, qualitative interviews with programme managers and experts were carried out.

This paper is structured as follows: First, we develop a typology of indicators and propose five criteria that disbursement-linked indicators should meet. Second, we classify and analyse the indicators used in five ongoing results-based programmes in the education sector based on the developed criteria. Finally, based on the findings of the study, policy recommendations are given on how indicators should be selected and defined.

2 Classification of indicators in results-based approaches

This section develops a classification of indicators used in results-based approaches. In general terms, an indicator can be defined as

> *a parameter, or a value derived from parameters, which points to/provides information about/describes the state of a phenomenon/environment/area with a significance extending beyond that directly associated with a parameter value* (Organisation for Economic Co-operation and Development [OECD], 1993, p. 5).[2]

2 In the context of development cooperation, the OECD/DAC (2009, p. 32) defines an indicator as a *"quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor"*.

Table 1: Typology of indicators

	Type of indicator	Definition	Illustrative examples and sample indicators
1)	Process indicator	Measure inputs used (i.e. financial, human and material resources) and actions taken or work performed to achieve a result	• Funds for primary education (input) • Number of workshops conducted (activity)
	Results indicator	Measure the short-term, medium-term and long-term results of a development intervention (i.e. outputs, outcomes and impacts)	• Number of classrooms built (output) • Students' learning outcomes (outcome) • Economic growth (impact)
2)	Direct indicator	Refers directly to the phenomenon/result it has been developed for	• A measure of standardised student test scores is a direct indicator of student learning outcomes
	Indirect indicator	Measures something (slightly or very) different from the phenomenon/result itself, nevertheless thought to paint a reasonably good picture of it	• Percentage of trained teachers is an indirect measure of education quality
3)	Quantitative indicator	Measures change in terms of numerical values that are objective or independently verifiable	• Number: Number of teachers trained • Percentage: Percent of government budget devoted to education sector • Ratio: Ratio of female-to-male school enrolment
	Qualitative indicator	Reflects subjective descriptions or categories	• Existence (yes/no): Primary school curriculum developed/not developed • Category: Level of parent satisfaction with school quality is 'high', 'medium' or 'low'

Sources: Own compilation based on Binnendijk (2000, p. 28); Danida (2006, p. 11); MDF Training & Consultancy (2005, p. 4); OECD (2008, p. 13)

This definition implies that indicators provide information that extends beyond the properties directly associated with a parameter value. Indicators provide information in summary form and reduce the information need that would normally be required to paint a precise picture of a situation (Delorme & Chatelain, 2011, p. 8; OECD, 1993, p. 5). Indicators do not, however, explain why a situation has arisen or a change has occurred (United Nations Development Programme [UNDP], 2002) and thus in themselves may not meet the strict scientific demands of demonstrating causal chains (OECD, 1993, p. 5).

Indicators are used in all types of development programmes to manage the implementation process and to report on results. In results-based approaches, indicators are specifically used as a basis to decide on the amount of funding to disburse.[3] While all development interventions require some form of evidence to disburse aid (e.g. documentary evidence of expenditures), the novelty in results-based approaches is that payments are directly linked

3 There are also several traditional, input-based projects and programmes that include a results component and use disbursement-linked indicators in addition to regular disbursement conditions and processes (see for example World Bank, 2012, p. 15).

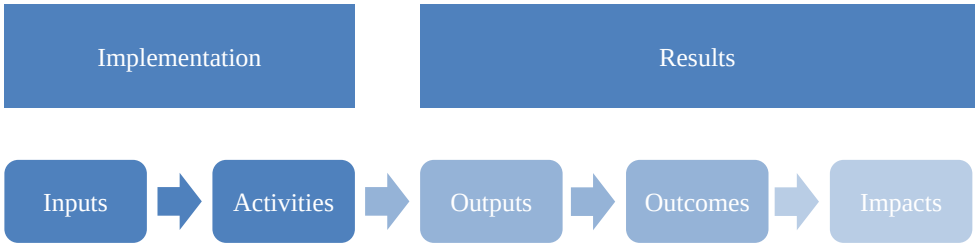
to the achievement of specific, pre-determined objectives or results which are measured through disbursement-linked indicators (DLIs)[4] (Gelb & Hashmi, 2014, p. 2). DLIs are often a subset of the indicators used for monitoring purposes and reflect inputs, activities and results which are considered as especially important towards achieving the overall programme objectives.

DLIs can be classified along three main categories (see Table 1) determining the level of results (outputs, outcomes and impacts or the inputs and activities necessary to strengthen programme performance); 2) measuring change in a direct or indirect way; 3) measuring change in a qualitative or quantitative manner.

2.1 Indicators at different levels of the results chain

DLIs can be defined at the different levels of the results chain (see Figure 1). The results chain is a logical and sequential model of steps that need to be taken in order to achieve the desired objectives. It starts with inputs, such as the amount of funding allocated to an intervention, followed by the activities undertaken to achieve the desired goals. The resources invested and actions taken contribute to results which are the describable or measurable changes derived from a cause-and-effect relationship (United Nations Development Group [UNDG], 2011, p. 15). There are three types of results of development interventions: the direct outputs, the short- and medium-term outcomes and the long-term impacts, which can be intended or unintended, positive and/or negative (OECD/DAC, 2009, p. 42).

Figure 1: The results chain



Source: Own illustration based on UNDG (2011, p. 14)

According to the OECD/DAC terminology, the different types of indicator used at the five levels of the results chain may be defined as follows:[5]

- **Input indicators**[6] measure financial, human and material resources used. Example: *The budget allocated to a vocational education programme.*

4 The focus on a 'disbursement-linked indicator' is specific to results-based approaches, but directly builds on previous experience with performance indicators that were used in budget support operations to guide aid allocation (Adam & Gunning, 2002; Koeberle, Stavresk, & Walliser, 2006).

5 The definitions are based on the Glossary of Key Terms in Evaluation and Results Based Management published by the Organisation for Economic Co-operation and Development/Development Assistance Committee (OECD/DAC, 2009).

- **Activity indicators** measure the actions taken or work performed as a result of which inputs such as funds, technical assistance and other resources are mobilised to produce specific outputs. Example: *Number of teacher training workshops conducted.*
- **Output indicators** measure the products, capital goods and services which result from a development intervention. Example: *Number of teachers trained.*
- **Outcome indicators** measure the likely or achieved short-term and medium-term effects of an intervention's outputs. Example: *Improvements in student learning outcomes among beneficiaries.*
- **Impact indicators** measure the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. Example: *Unemployment rates for youth aged 15-24.*

It is often difficult to classify indicators at the level of the results chain because it depends on the perspective of what an input, activity, output, outcome or impact is (ADB, 2013b, p. 19; Clist & Verschoor, 2014, p. 21). Generally, the results chain can be seen as a continuum where, at any point of the chain, the activity or result is an outcome of the previous process and an input to the next point in the chain (O'Brien, Fiszbein, Gelb, Kanbur, & Newman, 2012, p. 9). For example, from the perspective of programme beneficiaries, 'classrooms built' or 'teachers trained' may be seen as an input to increase the quality of education. From the perspective of a government or a development agency, on the other hand, both indicators represent common outputs of education interventions (ADB, 2013b, p. 19). Even an input indicator, such as the budget allocated to the education sector, could be seen as an output or outcome indicator if a budget increase is the intended result of a programme.

2.2 Direct and indirect indicators

DLIs can measure changes in a direct or indirect way. Direct indicators *"refer directly to the subject they have been developed for"* (MDF Training & Consultancy, 2005, p. 4). For example, an indicator defined as 'the number of children vaccinated' is a direct measure of the output of a child vaccination programme.

Indirect or proxy indicators measure *"something (slightly or very) different from the result itself, nevertheless thought to paint a reasonably good picture of the degree to which the result has been achieved"* (Danida, 2006, p. 11). Proxy indicators are often used when collecting data on direct indicators is difficult, unreasonably expensive or not feasible, either because the result itself is not measurable or because a change takes too long to unfold and cannot be measured in the short-term.

For example, a wealth index of household assets and housing characteristics is often used as an indirect measure of a household's living standard. Indexes can be particularly useful, when more direct measures of living standards such as income, expenditure and

6 The OECD/DAC (2009, p.32) defines inputs as the *"financial, human, and material resources used for the development intervention"*. In this paper, we use the term 'inputs' in a broader sense, i.e. not restricted to development interventions. For example, inputs could also be the resources used for a national or sector budget.

consumption are difficult and expensive to collect in developing countries (O'Donnell, Doorslaer, Wagstaff, & Lindelow, 2008, pp. 69-72).

In results-based approaches, process indicators (input and activity indicators) are often used as proxies when intended results are difficult to measure or take long to unfold. For instance, one of the intended results of the ADB's "Education Sector Development Programme" in Sri Lanka is *"institutional capacity at Ministry of Education and provincial levels strengthened"* (ADB, 2013a, p. 3). Since it is difficult to measure changes in the level of capacities, the programme uses several DLIs at input and activity level as proxies, assuming that measured inputs and activities are crucial for achieving the desired objective.[7]

2.3 Quantitative and qualitative indicators

Another distinction can be made between quantitative and qualitative indicators. Quantitative indicators measure results in terms of numerical values that are objective or independently verifiable, such as absolute values, percentages, rates and ratios (Binnendijk, 2000, p. 28; UNDP, 2009, p. 63; World Bank, 1996, p. 16). Some examples for quantitative indicators are:

- number of teachers trained;
- people with access to improved sanitation;
- proportion of population living on less than USD 1.25 a day;
- under-five mortality rate per 1,000 live births;
- poverty gap ratio at USD 1.25 a day.

Still, not all phenomena and results can be expressed in numerical terms. For example, the outcomes of interventions fostering democracy, good governance or institutional capacity-building are often qualitative in nature and hence can usually better be measured by qualitative indicators (Binnendijk, 2000, p. 29). Qualitative indicators can be

> *subjective descriptions or categories, such as whether or not a law has been passed or an institution has been established; beneficiaries' assessment of whether a project's services are excellent, satisfactory or poor; or simply a narrative describing change* (Binnendijk, 2000, p. 28).

Qualitative indicators can be expressed by nominal or ordinal variables. Nominal variables do not have a natural ordering, but are mutually exclusive (e.g. local anti-corruption law passed/not passed). Ordinal variables have a natural order, but the distance between values cannot be quantified (e.g. satisfaction with a job-training programme is low, medium or high) (University of California, Los Angeles [UCLA], 2013).

---

7 Among the indicators used are 'Sector Monitoring and Technical Support Unit (SMTSU) established with Program Head and key staff assigned' and 'Performance-based partnership agreements agreed and signed among Ministry of Education (MOE) and at least 7 provinces to enable MOE and provinces to align the provincial and central work plans and budgets for Education Sector Development Framework and Programme and which must confirm that at least 80% of the initially approved capital budget for school education in FY2013 has been spent' (ADB, 2013a, p. 3; 2013b).

3 Criteria for assessing indicators used to disburse payments

This section develops criteria for assessing the quality of disbursement-linked indicators used in results-based approaches. While recognising that there are numerous criteria that can be used to assess the quality of indicators, we only focus on those that are especially important to and in some cases specific to results-based approaches. To develop the criteria, we draw on literature that discusses criteria to select indicators for performance monitoring in traditional input-based interventions (see for example Delorme and Chatelain (2011); Holzapfel (2014)) and on literature dealing with results-based approaches (Clist & Verschoor, 2014; Kapur & Whittle, 2010; Savedoff & Martel, 2011). The experiences of the aid and development effectiveness agenda are also taken into account, in order to assess whether a programme is likely to perform well in terms of effectiveness and sustainability. The key criteria used in this paper are (1) focus on results, (2) control, (3) financial incentives, (4) measurability and verifiability, and (5) unintended consequences (see Table 2).

Table 2: Criteria to assess the quality of disbursement-linked indicators

Criterion	Key question	Considerations for results-based approaches
1) Focus on results	Do indicators ensure a focus on results?	• The indicators can measure results (outputs and outcomes) or processes (inputs and activities)
2) Control	Can results be influenced by and plausibly associated with the intervention?	• The extent to which incentivised actors have control over achieving the intended results • The extent to which results can be attributed to the intervention • The institutional setting of incentivised actors
3) Financial incentives	Can intended effects be maximised?	• The extent to which financial amounts reflect 'value for money', policy leverage, risk or other considerations • Whether disbursement is scaled in proportion to performance or conditional on achieving a threshold level
4) Measurability and verifiability	Are indicators reliable, consistent over time and independently verified?	• The relationship between the indicator and the underlying objective of the programme • The data quality and source (administrative data or survey data) • The way verification is organised (independent or not)
5) Unintended consequences	Can unintended effects be minimised?	• The extent to which indicators allow gaming (active manipulation of the indicators) • The extent to which indicators lead to distortions (indirect consequences of overemphasising or neglecting policy choices)

Sources: Authors

3.1 Focus on results

An increased focus on results is one of the main aims of results-based approaches and can be achieved by using DLIs at outcome level. The use of outcome-level indicators to make disbursement decisions has different advantages. Often, there is agreement between

development partners on desired development outcomes, such as the Millennium Development Goals, but uncertainty as to the causal chain or disagreement between the funder and the recipient on how the result can be achieved (O'Brien & Kanbur, 2013, p. 11). In this case, choosing a DLI at higher levels of the results chain, preferably at outcome level, is advisable.[8] It leaves complete discretion to the recipient on what inputs, activities and outputs to choose to deliver the agreed result and thereby increases ownership on the side of the recipient (O'Brien & Kanbur, 2013, p. 5). For instance, if an indicator measuring student test scores is chosen as a DLI, it is up to the recipient whether to improve teacher qualifications or to decrease the student-teacher ratio for instance. At the same time, the funder's aims are also satisfied because disbursements are only made once agreed results are delivered.

There are, however, cases where it may make sense to use DLIs at lower levels of the results chain. One reason is the need for pre-financing. The recipient in a results-based approach may not have the financial resources that are necessary to achieve the agreed final result of a programme. Indicators that measure inputs, activities and outputs allow for a more gradual and earlier disbursement of funds. Another reason is increased project supervision and the transmission of technical expertise from the funder to the recipient through the provision of a framework of necessary steps towards achieving the objective (O'Brien & Kanbur, 2013, p. 12). DLIs that are used in the World Bank's Program for Results or in a results-based lending programme of the ADB can for example be activities or processes necessary to strengthen programme performance (e.g. actions to strengthen the programme's M&E system or to improve fiduciary risk management) that are not directly related to desired overall programme results (ADB, 2013b, p. 20; World Bank, 2012, p. 35). However, the use of process indicators strongly limits the flexibility and discretion of the recipient in programme implementation. O'Brien and Kanbur (2013, p. 5) therefore suggest that DLIs at lower levels of the results chain should only be used, if there is certainty about the causal chain and if there is agreement between the recipient and the funder on what is needed to achieve the result. Otherwise, alternative means of pre-financing results, such as credits, may be more appropriate.

3.2 Control

The incentivised actor needs to have plausible control over achieving the intended results. Ideally, when payments are linked to indicators, these indicators should measure changes that are directly attributable[9] to the efforts of the incentivised actor(s). It should at least be possible to demonstrate that the efforts undertaken are one of the main causes of observed change. If the agreed result can only be influenced by the actor to a limited extent, two unintended scenarios may occur: In the first scenario, the actor makes strong efforts to improve the agreed measure, but there are external factors that negatively influence the result and the efforts of the agent are not sufficiently rewarded. In the second scenario, the

8 Impact indicators are not discussed here because they are hardly used as DLIs. Impacts are usually only observed after a development intervention has ended and are influenced by a multitude of factors which makes attribution of results difficult.

9 Attribution can be defined as *"the extent to which observed development effects can be attributed to a specific intervention or to the performance of one or more partners, taking account of other interventions, (anticipated or unanticipated) confounding factors, or external shocks"* (OECD/DAC, 2009, p. 21).

actor is rewarded, although no additional efforts have been undertaken and the result has only been achieved because of positive external influences (Clist & Verschoor, 2014, p. 7).

There is a trade-off between the results orientation of a results-based approach and the attribution of changes. Up to the level of outputs, it is fairly easy to attribute changes to specific interventions or development actors because inputs, activities and outputs are controlled by the stakeholders performing the intervention. Similarly, short-term outcomes – which are the changes arising directly from the output of an intervention (such as improved access to clean water) – are fairly easy to attribute. By contrast, medium-term outcomes and long-term development impacts, which are the ultimate aim of a programme, are influenced by many external factors, such as the economic and political environment and the activities of other development agencies. These factors make attribution difficult (Prennushi, Rubio, & Subbarao, 2002, pp. 108-110)

When selecting indicators, it must also be kept in mind that actors – whether at the level of individuals, service providers, line ministries or central government – always operate within existing local incentive systems, institutional structures and interdependences. For instance, an education ministry might depend on the support of other line ministries (transport, rural development, finance) or sub-national entities (local and regional governments) for achieving the agreed education results. Such relations therefore need to be considered when selecting DLIs. Funders and partner countries should therefore work closely together from the start and follow a 'line of sight' approach to align the goals of crucial actors and institutions with the goals of the results-based approach.

3.3 Financial incentives

Financial incentives attached to indicators are the core elements of results-based approaches. Different considerations have to be made with respect to financial incentives. First, pricing results and allocating specific amounts for each indicator have to be decided. Second, it has to be determined whether a target is set for an indicator or whether incremental payments will be made for each additional unit of progress.

Compared to traditional aid modalities that base disbursements solely on inputs, a results-based approach prices outputs and outcomes. For pricing results, no clear method exists and different ways are suggested to decide on the amount of financing to allocate for individual DLIs. Among the most important criteria are 1) value for money, 2) potential leverage effects, and 3) additional risks for partners (O'Brien & Kanbur, 2013, pp. 15-16; Savedoff & Martel, 2011, pp. 5-6).

From a 'value for money' perspective, it is important for funders that financing allocated to a DLI is reasonable, i.e. that payments do not exceed the estimated cost to achieve the desired results, also in relation to similar donor-financed activities. But Savedoff and Martel (2011, p. 6), argue that the amount of payment should not primarily be based on value for money considerations, but rather on the amount needed to attract the attention of policymakers (leverage). Due to the often high amounts of development aid in sectors such as health or education, relatively high payments may be needed to induce incentives that are large enough to raise the attention of policymakers.

From a 'risk perspective', partner countries could also expect greater ex-post financial rewards for bearing more risk in comparison to traditional input-financed activities that disburse all funds upfront (Energy Sector Management Assistance Program [ESMAP], 2013). According to Clist and Verschoor (2014, pp. 7-8), the amount of the risk premium is influenced by two factors. The first is the level of risk aversion: a more risk-averse actor would expect a greater risk premium. The second is the level of control of the actor has over achieving the result. If the actor only has limited control over achieving a DLI, the agent may only agree to a results-based approach that adequately compensates for the higher risk of not achieving the agreed result.

Usually more than one DLI is used. This leads to the question of how much funding should be allocated to each indicator, and how the share of funding per indicator influences incentive structures. If equal funding is allocated to each indicator, it is likely that attention is focused on the complete results framework and a holistic approach is taken to achieve all dimensions. Yet, there is the danger that such an approach leads to a focus on those results that are easiest to achieve. It may therefore be preferable to allocate a higher share of funding to those indicators that are more difficult to achieve. Another approach could be to put a higher weight in terms of payments on those indicators considered to be especially important to the involved actors (O'Brien & Kanbur, 2013, p. 16).

The second question with regard to financial incentives is whether disbursement is scaled in proportion to performance or conditional on achieving a threshold level (Gelb & Hashmi, 2014, p. 12). When payment rewards incremental progress, relatively stable incentives are provided. It does, however, become more difficult to deliver results once certain progress has been made. This could be reflected by successively increasing the amount of payment per unit.[10] Such a phased approach for successively adjusting payments per unit can help to avoid performance plateaus. In comparison to scale-based payments, DLIs that are threshold-based provide high incentives for performance at the margin but low incentives if the threshold is set too low or too high (i.e. the threshold can be achieved with minimum effort or achievement seems unlikely) (Gelb & Hashmi, 2014, p. 15). Setting targets is challenging because it is difficult to make a realistic assessment of what can be achieved with reasonable efforts and available resources within a given period.

If payments are made for each unit of incremental progress, it has to be decided whether total achievements (e.g. every child who survives to age five) or additional achievements (e.g. every child who survives to age five above a baseline) are rewarded (Savedoff & Martel, 2011, p. 6). Also, baselines can be static (fixed at the beginning of the programme) or dynamic (adjusted every year). Savedoff and Martel (2011, p. 7) argue that paying for additional improvements is preferable because it creates a higher incentive for the recipient to improve over current conditions. Paying for total achievements implies that payments are even made when results decrease from year to year. However, paying for additional improvements is only possible if adequate baselines exist.

---

10 For example, one of the recommendations of the evaluation of the first five years of GAVI immunisation service support is as follows: *"GAVI should consider additional and/or different measures of immunization performance in higher coverage countries – such as improving equity or coverage consistency. GAVI's focus on the number of additional children immunized becomes less appropriate in higher coverage countries, as costs of increasing coverage are harder to justify in terms of disease reduction, and the amount of reward funding that countries will receive becomes lower as coverage increases and it becomes harder to immunize additional children."* (Chee, Hsi, Carlson, Chankova, & Taylor, 2007, p. xvi).

When incremental progress is rewarded and targets are not set, another question is whether annual financial ceilings should be used. From a funder's point of view, financial ceilings are necessary to avoid payments exceeding the programme budget. It has, however, been argued that ceilings can have a negative impact on incentives because they act as an implicit target. If only a low share of the available amount is disbursed for a given year, the perception can arise that the programme is a failure (Birdsall, 2013). Besides, undisbursed funding in one year often cannot be reallocated to subsequent years for reasons of accounting and might therefore be 'lost'. Further, annual ceilings may discourage the incentivised actor to undertake efforts that only yield results in later years. The actor might worry that the funder is not able to adequately reward performance in the future (Birdsall & Perakis, 2012, pp. 3-4).

3.4 Measurability and verifiability

When disbursements are based on results, it is important that those results are measurable and independently verifiable to ensure credibility.

First, indicators should capture the efforts or results that are to be measured as best as possible. Generally, direct indicators should be preferred. Proxy indicators should only be used, if it is sufficiently clear that there is a strong correlation between the proxy indicator and the subject of interest (United States Agency for International Development [USAID], 2010, p. 5).[11] In this regard it is important that there is alignment between the actions required to improve the value of the chosen (proxy) indicator and those actions required to achieve the underlying objective (Clist & Verschoor, 2014, p. 21). If this is not the case, misdirected incentives may lead to undesirable actions and unintended effects.

Second, indicators used in results-based approaches have to be consistent and reliable over time, available on a timely basis, and sensitive to changes (i.e. the indicator value should change if results change).

Third, data has to be credible and independently verifiable. For funders to engage in a results-based approach, it is important to know that they only pay for 'true' results. Incentivised actors on the other hand need to be convinced that measurement and compensation of their efforts are fair and transparent.

While data quality and credibility is crucial, the costs of data collection, analysis and verification have to be kept within reasonable limits. Various problems arise with regard to data quality and verifiability at the different levels of the results chain.

Data on indicators measuring inputs, activities and outputs are easy to collect and can usually be drawn from financial accounts, programme records or as administrative data systems from sector ministries. Further, it is relatively easy and cost-effective to verify these types of data. Yet, there can be room for interpretation when verifying whether certain activities or outputs have been delivered as agreed. For instance, an indicator measuring

11 Indicators should also be direct in a sense that they reflect the measured objective at the right level of the results chain, i.e. an objective at outcome-level should ideally be measured by an outcome indicator (USAID, 2010, p. 5).

whether a reform was implemented as planned may be amenable to interpretation, especially when quality aspects have to be considered as part of the verification process.

Measuring development outcomes is even more challenging. One option is to rely on country-level monitoring and evaluation (M&E) systems to measure progress. Data on many development outcomes (e.g. enrolment rates, data on student-teacher ratio) can be obtained from government administrative systems. While administrative data is cheap to exploit, there are frequent problems with the accuracy, timeliness and comprehensiveness of administrative data (Pearson, Johnson and Ellison, 2010, p. 28; Prennushi et al., 2002, p. 113). Another option for measuring development outcomes is to use household survey data. The risk of data manipulation is lower when survey data is used because households have fewer incentives to report incorrect data. Programme administrators or local officials in contrast might be more inclined to manipulate data, especially when budget allocations or rewards depend on the data reported (Prennushi et al., 2002, p. 114). However, data obtained from government household surveys is often unreliable and suffers from sampling error and biases (Pearson et al., 2010, p. 4). In addition, most household surveys are not carried out annually, as would be required in most results-based approaches (Pearson et al., 2010, p. 30).

Often, the best option with regard to data quality and reliability is to create parallel administrative systems or to carry out independent household surveys only for the purpose of a results-based approach. Yet, such an approach is costly and M&E systems created might not be useful after a specific results-based programme ends (Pearson et al., 2010, p. 32). Besides, the use of parallel M&E systems is not in line with the aid effectiveness agenda. International donors have committed to adopting country-level results frameworks to assess progress and to minimise their use of additional frameworks and indicators (Busan Partnership, 2011). To ensure data quality and to rely on national statistical systems simultaneously, it might be advisable to use existing survey data or administrative data (if available) coupled with independently verification. For example, data could be verified by an independent household survey (Birdsall, Savedoff, Mahgoub, & Vyborny, 2011, p. 5) or by randomly checking administrative data (Perakis & Savedoff, 2015). To verify enrolment data, for instance, a random number of households could be visited to confirm that children reported as enrolled actually go to school. If government data is used, it may often be necessary to provide capacity development support to strengthen national statistical systems (Keijzer & Janus, 2014).

3.5 Unintended consequences

Linking payments to the achievement of measurable results may induce incentives for gaming and lead to distortions (Clist & Verschoor, 2014, pp. 13-14). Gaming in this context is defined as ‘reactive subversion’ of a results-based approach (Bevan & Hood 2006, p. 521) while distortions are understood as indirect consequences of a focus on specific indicators. Generally, the higher the potential reward, the higher the risk of unintended effects (Perakis & Savedoff, 2015).

One example of gaming is measure fixation, i.e. an “*emphasis on measures of success rather than the underlying objective”* (Smith, 1995, p. 290). If a DLI does not capture all the aspects of an underlying objective, the recipient is encouraged to pursue strategies that contribute to

raising the value of the DLI rather than strategies that help to achieve the associated objective. Clist and Verschoor (2014, p. 11) give the example of a results-based approach with the underlying objective of creating an educated and productive work force, in which the indicator 'number of students taking a final exam' is used to reward performance. The incentivised actor can chose between different actions to increase the completion rate. One option is to improve the quality of education, which would also contribute to the underlying objective. Another option is to discourage drop-out or repetition, even in cases where it would have a positive effect on a student's learning outcomes.

Data manipulation by the incentivised actor is another form of gaming. It may become a problem, if the recipient has a role in collecting and monitoring data on DLIs (e.g. through self-reported administrative data), and when scope for external verification of results is limited. For instance Sandefur and Glassman (2014, p. 13) show that the incentive introduced by the Global Alliance for Vaccines and Immunisation (GAVI), which offered to pay USD 20 per additional child immunised against diphtheria-tetanus-pertussis (DTP3), led governments to over-report DTP3 coverage by 5%. In the education sector, the use of examination data to assess performance is seen critically. Although examination results can be a good indicator for learning outcomes, they also entail a high risk of gaming. When examination results are used to pay for performance, tests become 'high-stakes' tests[12], where actors have a strong incentive that students perform well (Lockheed, 2008, p. 13). This pressure to perform may lead to different adverse effects, such as 'teaching to the test', the rejection of weaker and hard-to-teach students from the system, and ultimately inflated test scores without actual improvements in learning (Lockheed, 2008, p. 13; Shepard, 1991). The risk that students, teachers or schools cheat is particularly pronounced, if performance incentives are directly paid to schools or teachers (Lockheed, 2008, p. 13).

Longer contract periods can reduce the risk of gaming to some extent. The shorter the contract period, the greater the scepticism on the side of the actor about whether genuine efforts will lead to desired results within the contract period (Clist & Verschoor, 2014, p. 21). When planning a results-based approach, it is therefore important to take into account that efforts and investments, in particular if they are aimed at longer-term changes, often take years to be reflected in outcome measures.

Distortions may occur as a more indirect consequence of linking payments to results. Results-based approaches incentivise actors to concentrate efforts on results associated with DLIs, which may lead to a neglect of other important goals (Birdsall et al., 2011, pp. 32-33). For instance, paying for improvements in the enrolment rate could lead to a neglect of the quality of education.[13] This type of unintended effect can partially be mitigated by using several related DLIs, which pay for improvements in quality and quantity (Pearson et al., 2010, p. 28). However, a complex set of indicators may diffuse

12 In comparison, 'low-stakes' tests are not used to make decisions about individuals or groups although they can be used to make a comparison between students to inform decision-making (Lockheed, 2008, p. 13).

13 The GFATM (The Global Fund to Fight AIDS, Tuberculosis and Malaria) evaluation, for example, found that *"basing the GFATM's PBF (performance based funding) system largely on numeric output targets created unintended negative consequences, especially in terms of the quality of service provision. Implementers in more than half the SA2 countries reported that, on at least one occasion, they had sacrificed quality of implementation in order to achieve a quantitative numerical PBF output target"* (Pearson et al., 2010, p. 36).

the attention of responsible actors and may hence be less effective in generating incentives (Birdsall et al., 2011, p. 64). Another possibility to mitigate such unintended effects is to strengthen and to use a country's public accountability mechanism. Birdsall et al. (2011, p. 58) propose that when using the DLI 'number of students who take a test' in a COD Aid agreement, it should be required that test scores are publicly disseminated. This could add important non-financial incentives in the form of peer pressure or public naming and shaming to the financial incentive.

A different form of distortion is 'cherry picking'. Incentivised actors may concentrate on helping those closest to the performance threshold, where least efforts have to be exerted to achieve the target (Pearson et al., 2010, p. 75). For example, if payments are made to teachers for each additional student passing an exam, teachers may focus only on those students who are just below the threshold and neglect both the weakest students and those who already perform well. Paying for results may thus lead to efforts being concentrated on better-off groups that are easy to reach and on regions where results are easy to produce (Birdsall et al., 2011, pp. 32-33). One option to reduce the risk of such unintended effects is to disaggregate DLIs and to make higher payments for reaching disadvantaged groups (e.g. women, ethnic groups) or remote regions.

4 Overview of results-based approaches in education

Various funders have started implementing results-based pilot programmes in the education sector. This chapter classifies and assesses the disbursement-linked indicators used in five pilot programmes based on the typology and criteria developed in Sections 2 and 3. The five programmes are:

- World Bank "Big Results Now in Education Program" Tanzania
- ADB "Education Sector Development Program" Sri Lanka
- DFID "Results Based Aid Pilot – Education Sector Programme" Rwanda
- DFID "Pilot Project of Results Based Aid (RBA) in the Education Sector" Ethiopia
- Twaweza "KiuFunza Local Cash on Delivery" Tanzania

These pilot programmes were selected as case studies because they have similar objectives (improved access to and quality of education; improved learning outcomes) and represent innovative forms of introducing a greater focus on results in development interventions. We selected two pilot programmes implemented by multilateral development banks, an ADB results-based lending pilot and a World Bank PforR pilot, and three pilot programmes that are based on the COD Aid model. Comparing these different types of programmes gives an insight into current applications of results-based approaches in the education sector. The education sector is already far advanced in implementing results-based approaches and the selected pilots offer a cross-country perspective. The findings from the individual assessments and the comparison across these pilots have validity for the specific context of each intervention as well as for the discussion on results-based approaches in development cooperation as such. Many of the findings can be translated to

other results-based approaches that are implemented by governments and NGOs, also beyond the education sector.

Table 3 provides an overview of the five programmes selected. The programmes of the development banks are relatively complex: they have several objectives, a comparatively large number of DLIs (16 for Tanzania and 21 for Sri Lanka), longer contract periods (5 years), and large budgets that are disbursed as loans (USD 122 million in Tanzania and USD 200 million in Sri Lanka). Payments in both programmes go to the partner governments. While financing is not linked to individual transactions within a programme, disbursements should not exceed total programme expenditures (ADB, 2013b, p. 28; World Bank, 2012, p. 40).[14]

In comparison, the design of the programmes based on the COD Aid model has been kept simple. The programmes are characterised by a smaller number of DLIs (4 in Rwanda, 2 in Ethiopia and 3 in Tanzania) and shorter contract periods (3 years). Two of the programmes (Rwanda and Ethiopia) are implemented by DFID. The budgets are comparatively small (USD 14 million in Rwanda, USD 47 million in Ethiopia) and disbursed as grants. The recipient in both cases is the Ministry of Education and funds are earmarked to the education sector. The third programme KiuFunza ('thirst to learn') in Tanzania is implemented by a local NGO, Twaweza, and makes payments directly to teachers. Although this pilot does not comply with the conventional definition of a results-based aid programme (contract between governments), it is included here because the setup mirrors the definition of COD Aid at a local level (the programme documents call the design 'local cash on delivery'). The amount of funding (approximately USD 585 thousand) is low because the programme is implemented on a pilot basis in only 154 schools to explore the possibility of scaling up nation-wide.

The quality of indicators is assessed in relation to the overall objectives and results statements[15] of the five programmes.[16] All results-based approaches follow a specific 'theory of change', a sequence of steps that is expected to lead to a particular desired outcome (Vogel, 2012). There might be different intermediate objectives as part of the theory of change that are considered as important to achieve the final objective. The programmes of the development banks define results statements for intermediate and final objectives and use several clusters of DLIs to measure progress towards each intermediate result. In comparison, the programmes based on the COD Aid approach only use one set of DLIs to measure the achievement of the overall objective.

14 If disbursements by the World Bank together with contributions to the programme from other financing sources exceed the amount of programme expenditures, the recipient has to pay back the difference to the World Bank (World Bank 2012, p. 40).

15 The quality of results statements is a precondition for the selection of appropriate indicators and can be tested against the SMART criteria. These are: specific, measurable, achievable, realistic and time-bound. The SMART goals were originally developed by Doran, Miller and Cunningham (1981).

16 In this context, it has to be mentioned that results-based approaches often have several objectives and not all are measured by disbursement-linked indicators. For example, besides improving the quality of education, a donor may aim to use a results-based approach to encourage a stronger results orientation or 'results culture' on the partner side or aim to strengthen the aid relationship between donor and recipient (see for example Department for International Development [DFID], 2011b).

Table 3: Results-based approaches in education

Funder and approach	Partner	Overall objective	No. of DLIs	Total amount of financing (in million USD)	Contract period (years)
Multilateral development banks					
World Bank Program for results	United Republic of Tanzania	Improve the quality of basic education service delivery, thereby producing tangible improvement in learning outcomes	16	122	5
ADB Results-based lending	Democratic Socialist Republic of Sri Lanka	Development of a modernised secondary education school system with the overall objective of increasing youth employability	21	200	5
Approaches based on the cash on delivery aid concept					
DFID Payment by results (RBA)	Ministry of Education (Rwanda)	Equitable access to quality education and improved learning outcomes	4	14[a]	3
DFID Payment by results (RBA)	Ministry of Education (Ethiopia)	Improved access to, and quality of lower secondary schooling.	2	47[b]	3
Twaweza Local cash on delivery	Teachers (Tanzania)	Improve early grade learning outcomes	3	0.59	3

[a] GBP 9 million (exchange rate of 15 December 2014: GBP 1 = USD 1.57)
[b] GBP 30 million (exchange rate of 15 December 2014: GBP 1 = USD 1.57)

Sources: Authors

4.1 Analysis of the indicators used in five selected programmes

4.1.1 World Bank "Big Results Now in Education Program" Tanzania

The World Bank funds the "Big Results Now in Education (BRNEd) Program" as the first Program for Results pilot in the education sector. A distinctive feature of this programme is a close link and alignment with a national reform agenda, namely the Big Results Now initiative by the Tanzanian Government. This initiative focuses on six priority areas of the national administration, including education. Big Results Now (BRN) aims to deliver key measurable results within a specified timeframe, and sector plans were developed in intensive eight week 'delivery labs' for each sector (DFID, 2013a). The model closely follows the example of Big Fast Results Now reforms in Malaysia which is implemented with the support of the Performance Management & Delivery Unit, a unit of the Malaysian Prime Minister's office (Aman & Balozi, 2014). A key characteristic of the Tanzanian BRN is that responsible ministers have taken public pledges to be held

accountable for achieving the agreed results that have been set. Overall, this initiative demonstrates a strong existing results orientation in the public administration in Tanzania and a keen interest of development partners to support such an approach.

The objective of the BRNEd is to improve the quality of basic education service delivery, thereby producing tangible improvement in learning outcomes and, in the longer term, to lay a foundation of an *"outcome-based performance culture in the education sector in Tanzania"* (World Bank, 2014, p. 13). The World Bank will contribute USD 122 million over five years (2014-2018) and a total of 16 DLIs have been set up (see Table 4).[17]

Table 4: World Bank "Big Results Now in Education Program" Tanzania

	Input	Activity	Output	Outcome	Total
Number of indicators	3	10	2	1	16
Amount of financing					
Incentivised actor	Government of Tanzania				
Million USD	43	43	20	16	122
% of financing	35.5%	35.5%	16%	13%	100%

Indicator typology			
Direct	13	Quantitative	7
Indirect	3	Qualitative	9

Notable characteristics					
No. of ministries	3	Independently verified indicators	5	Central-level indicators	5
No. of donors	3	Indicators verified by donors	11	Local-level indicators	11

Outcome indicator: 'The recipient has met the annual target of improvement in words per minute (wpm) in national 3R [reading, writing and arithmetic] average FY [Financial Year]'

Sources: Own compilation based on World Bank (2014)

The BRNEd Programme is relatively complex. The overall objective of improving education quality in primary and secondary schools has been operationalised through six intermediate objectives or clusters of DLI (see Table 5). In total, there are 16 DLIs and between one and five DLIS per intermediate objective.

17 In total, the budget of the BRN in the education programme is USD 416 million, with contributions from the World Bank (USD 122 million), the United Kingdom (UK) (USD 100 million), Sweden (USD 30 million) and the Government of Tanzania (USD 164 million). The analysis here will focus mostly on the World Bank contribution, as it represents the biggest share and has been most advanced in terms of implementation compared to the other donor contributions.

Table 5: The six intermediate objectives/clusters of DLIs

	Intermediate objectives / clusters of DLIs	Level of the results chain	No. of DLIs	Amount of financing (in million USD)	Verification by
1	Recipient has completed all the Foundational Activities	Activity	5	15	Independent firm
2	Recipient has evidenced timely and adequate resource flows for the Program	Input	2	38	Donors
3	The Recipient has produced an Annual Summary Education Performance Report (ASEPR) and annual school-level Education Management Information System (EMIS) data by each Local Government Authority (LGA)	Activity	2	12	Donors
4	Teachers have been deployed efficiently across and within districts	Output	2	20	Independent firm
5	Number of schools receiving school incentive grant (SIG) as indicated in the Program design	Activity / Input	3	18	Donors and independent firm
6	Recipient has demonstrated an increase in student learning outcomes	Outcome	2	19	Donors

Sources: Own compilation based on World Bank (2014)

Focus on results

The programme targets multiple points along the results chain and the DLIs are a mix of outcome, output and process indicators. The overall design of PforR is less focused on results at the outcome-level and heavily relies on process indicators. Most DLIs reward activities like formulating budget frameworks, producing reports or setting up other processes. The only outcome indicator rewards improvements in reading, writing and arithmetic assessments but the funding attached to this DLI only represents about 13% of the overall amount of funding. In addition, two indicators measure outputs and the amount of financing linked to these indicators represents 16% of total funding. The majority of funds (71%) are linked to either input indicators (three DLIs) or activity indicators (ten DLIs).

Disbursements per indicator vary in terms of volume and the timing of disbursements. Foundational activities such as agreement on budget framework (DLI 1.1) or preparing a list of all primary and secondary schools (DLI 1.4) are only measured in the first year and disbursements only take place once. In contrast, for other DLIs such as the number of schools achieving acceptable student-teacher ratios (DLI 4.2), disbursements take place annually. In general, the volume of disbursements goes down over the implementation period. For the first year, USD 51 million are allocated, for the second and third year USD 27 million and for the last year only USD 17 million.

The rationale of the declining amounts of funding might be that setting up processes at the beginning is most costly. Programme documents (World Bank, 2014, p. 3) state that the initial stages of implementation aim at *"system level improvements in form of more efficient utilisation of financial and human resources*", whereas at the middle of implementation

targets *"tangible improvements in actual education service delivery"* and the end of programme implementation focuses on *"tangible but realistic improvements in student learning outcomes"*. This gradual approach of increasing the ambition of the programme appears feasible as it raises the level of ambition from influencing outputs towards improving outcomes over the implementation period. However, the declining amounts of funding in the last year of implementation do not reflect this aspect, as one could expect that better results should also trigger higher rewards.

Control

The indicators are closely linked to activities that are within the control of government ministries and can easily be attributed. A complicating factor is that competencies for education policies are dispersed among different government ministries in Tanzania. Hence, the programme targets three ministries: the Ministry of Education and Vocational Training; the Prime Minister's Office (Regional Administration and Local Government); and the Ministry of Finance. In addition, the programme supports the Big Results Now initiative that has also set up additional management structures, including a central Presidential Delivery Bureau and individual Ministerial Delivery Units in sector ministries. All of these governmental bodies have different functions in releasing funds and in setting and coordinating policies in the education sector. The programme design closely tailors to the respective responsibilities and actions of each government body, and thus control of achieving the desired results should be high.

In addition to different ministries at the level of central government, the programme targets the local level by incentivising Local Government Authorities (LGAs). In total, eleven DLIs target the actions of central government ministries and five DLIs address the actions of local governments that are responsible for providing schools with funds and teachers. This multi-ministry and multi-level setup can be described as complex and requires strong coordination efforts. The advantage could be that incentivised parties have a high degree of control over achieving the DLIs. The disadvantage would but be that the overall setup is quite complicated and focuses more on processes than on outcomes.

Financial incentives

The disbursements are either based on achieving certain thresholds or based on units on a scale. Out of 16 DLIs, 9 are formulated as thresholds, mostly measuring whether budgets, plans or measurement tools have been set up. Disbursements for these DLIs are mostly made once in the first year and the indicators are qualitative indicators formulated in binary terms. As a consequence funding is either released in full or not at all.

The remaining seven DLIs are based on different scales where proportional progress is rewarded. For most of these DLIs, annually allocated disbursements are made (in quarterly payments) in a pro-rated manner for percentages of funds released or the number of schools and local governments meeting certain indicators. For instance, DLI 3.2 'recipient has made available online an annual school-level EMIS [Education Management Information System] data' is linked to USD 2 million annually for the first three years. The annual amount is divided equally among all LGAs in the country, and each LGA receives a pro-rated share in terms of percentage of schools for which data is available online.

For the DLIs that are based on scales, all reward per-unit progress, except for one DLI where in addition to per-scale payments an explicit target has been set: The outcome indicator DLI

6.2 ‘recipient has met the annual target of improvement in words per minute (wpm) in national 3R average FY [Financial Year]’. A quarter of the USD 16 million allocated to DLI 6.2 is disbursed per unit progress, whereas the remaining funds are disbursed for improvements in words per minute over the previous year, above a threshold of 0.5 wpm.

The incentive effect of scale indicators that set targets differs strongly from indicators where each unit of progress is rewarded. For instance, if the government slightly misses the target of DLI 6.2, for instance 0.4 wpm instead of 0.5, the annual disbursement of the budgeted USD 4 million does not take place, even though significant progress might have been made. Whether this target is achievable or whether it will have to be revised is still open. Another interesting feature of DLI 6.2 is that annual improvements are used, which means that the baseline is dynamic and changes every year. In comparison to a static baseline, a dynamic baseline sets a stronger incentive to achieve additional results every year.

Undisbursed funds can be reassigned to DLIs for the final programme year and an allocation of funds over the remaining period of programme implementation will be carried out at mid-term. These reallocations could potentially affect the outcome orientation of the overall programme and undermine existing incentive structures. For instance, the funders could lose credibility if they shift undisbursed funds between DLIs and reward the recipient, even though a number of results might not have been achieved.

Measurability and verifiability

Most DLIs are proxies for the overall objective of improving learning outcomes. Except for the outcome indicator, it is not clear that the incentivised resource flows and management activities will translate directly into improved education results at the outcome level. Most DLIs are direct and qualitative and assess the performance of government entities. While these indicators are direct in the sense of targeting specific activities, they are only an indirect measure for the overall objective.

The outcome indicator is based on the 3R (reading, writing, arithmetic) assessment, which is part of the Big Results Now reform, and promises to provide a better baseline for assessing student learning outcomes than existing tests (World Bank, 2014). Existing tests only assess pass rates after primary and secondary schools, and have strongly fluctuated over recent years and even declined heavily. The 3R assessment will be undertaken in a sample-based manner in primary schools, and is designed to diagnose learning achievement throughout the system, even before graduation.

The data that is used for DLIs is all administrative data that comes from the government’s own systems and mainly three sources. Data for DLI measurement will draw on basic education data from the national Education Management Information System, student performance data from the semi-autonomous National Examinations Council of Tanzania and specific dashboard data which has been developed for the Big Results Now initiative (overseen by the BRN Presidential Delivery Bureau). The government has also developed Big Results Now budget codes for incentive payments that go to local governments, in order to track expenditures (DFID, 2014a).

To support data collection and measurement, donors have targeted technical assistance at different levels in order to create a detailed online database with timely and comprehensive information at the school level. In addition, two indicators (DLI 3.1 and 3.2) incentivise

national and local government to be compliant by releasing annual national education performance reports and publishing school-level data online.

The verification of the 16 DLIs is semi-independent. Five DLIs will be independently verified by a private actor that will be identified through an open tendering process. The remaining DLIs will be assessed by the development partners (World Bank, DFID and Sweden). The independent verification will mostly focus on the local-level activities, whereas the development partners verify central-level indicators targeting the different ministries. This setup provides some incentives for improved performance at the local level, but verification by development partners poses the potential risk of skewing the power balance between development partners and the Government of Tanzania more than a completely independent verification process would.

Unintended consequences

Unintended consequences will have to be carefully monitored, but not enough evidence is available to draw conclusions yet. The semi-independent verification might present one potential risk of an unintended consequence related to disbursement pressures on the side of the donors. If results for donor-verified DLIs are not achieved, then the donor group might still feel tempted to disburse funds, and technically they would be free to make this decision. Or the government might feel unfairly treated by the assessment.

Another potential unintended consequence is related to the separate technical assistance package (USD 7 million over four years). On top of the PforR programme, DFID funds several technical advisors and capacity development support that will ensure smooth operation of the PforR (World Bank, 2014). However, such a setup bears the risk of 'instrumentalising' technical assistance to make disbursements under the PforR, without addressing genuine capacity needs. For instance, technical assistance will develop a *"Program Operations Manual to cover workflows and detailed processes for meeting DLI targets"* and raise *"awareness about the DLI linked incentives"* (World Bank, 2014, pp. 23-24). This could lead to a form of distortion, if technical assistance is geared towards triggering disbursements from donors only. In an extreme scenario, donors would pay for results that they themselves are responsible for achieving (Keijzer & Janus, 2014).

In line with other PforR programmes, an integrated risk assessment was carried out before implementation started. These assessments investigate technical, fiduciary, environmental and social, and DLI-related risks. Many potential unintended consequences and respective mitigation strategies have already been identified in advance and included in the programme design. For this programme, the risk was assessed as substantial. The main risks that were identified include country elections in 2015, weak management and the supervision capacity of implementing ministries and the relationship between Big Results Now new appointees and public servants of the present education system. Whether these risks will lead to unintended effects, or whether they can be mitigated, remains open.

In addition, the government plans to mitigate potential gaming risks around student assessments (for instance, artificial improvement of pass rates by disallowing weaker students to take the exams) but it is still identifying methods to minimise gaming of the exam results (World Bank, 2014). One option will be to adopt a variety of verification processes, which will test the validity and reliability of the results.

4.1.2 ADB "Education Sector Development Programme" Sri Lanka

The Education Sector Development Programme (ESDP) in Sri Lanka is the first Results-Based Lending (RBL) programme (USD 200 million loan) approved by the ADB board and supports the Sri Lankan government in the development of a modernised secondary education school system with the overall objective of increasing youth employability (ADB, 2013c, p. 3).[18] The programme is embedded in a sector-wide approach and contributes to the government's Education Sector Development Framework and Programme 2013-2017.[19] The Ministry of Education is responsible for executing the programme. Implementation is in the hands of the Ministry of Education and the nine provincial education departments (ADB, 2013c, p. 5). Table 6 provides an overview of the programme and classifies the indicators used.

Table 6: ADB "Education Sector Development Program" Sri Lanka[a]

	Input	Activity	Output	Outcome	Total
Number of indicators	1	9½	6½	4	21
Amount of financing					
Incentivised actor	Government of Sri Lanka				
Million USD	12	68	80	40	200
% of financing	6%	34%	40%	20%	100%

Indicator typology			
Direct	13½	Quantitative	9½
Indirect	7½	Qualitative	11½

Notable characteristics					
Contract period (years)	5	Independently verified indicators	3	Indicators to strengthen programme performance	6
		Indicators verified by government/donor	18	Indicators targeting the central and provincial level	2

Outcome indicators

- General Certificate of Education 'O' Levels pass rate
- General Certificate of Education 'A' Levels pass rate
- Enrolment in General Certificate of Education 'A' Levels Commerce Stream
- Enrolment in General Certificate of Education 'A' Levels Science Stream

[a] Some of the indicators are composite indicators and consist of two individual indicators. In some cases, individual indicators within a composite indicator have different characteristics (qualitative/quantitative, direct/indirect) or are defined at different levels of the results chain. If this is the case, each individual indicator is categorised separately and given the weight of half an indicator.

Sources: Own compilation based on ADB (2013c)

18 The expected results of the programme are: (1) improved student learning, (2) improved equity and efficiency of the school system, (3) strengthened school leadership, (4) strengthened capacity for effective programme planning and implementation (ADB 2013e, pp. 3-4).

19 See the Education Sector Development Framework and Programme (ESDFP) (2013 – 2017) (Ministry of Education Sri Lanka, 2013).

The ESDP is a relatively complex programme. Nine intermediate objectives[20] have been set and were identified as important steps towards increasing youth employability in the programme's results framework (see Table 7). Four of the intermediate objectives (increased enrolment and increased pass rates) are defined at outcome level and the remaining at output level. Progress towards each of the intermediate objectives is measured by a separate set of DLIs. In total, there are 21 DLIs and there is a maximum of one DLI per intermediate objective per year. Funds are disbursed equally over the 5-year programme frame, i.e. every year a maximum of USD 40 million can be disbursed depending on the progress made.

Table 7: Intermediate objectives of the ADB Education Sector Development Programme Sri Lanka

	Intermediate objectives	Level of the results chain	No. of DLIs	Amount of financing (in million USD)	Verification by
1	Pass rates for General Certificate of Education (GCE) 'O' level examinations increased	Outcome	1	10	Government (DOE)
2	Pass rates for GCE 'A' level examinations increased.	Outcome	1	10	Government (DOE)
3	Pathways from school to Technical and Vocational Education and Training (TVET) developed – Technology Stream commenced and implemented at GCE 'A' levels.	Output	5	32	Government (SMTSU/MOE), ADB (independent verification in Year 4)
4	Secondary schools upgraded to offer all subject streams	Output	4	32	Government (SMTSU/MOE), ADB (independent verification in Years 2, 3 and 4)
5	Enrolment in GCE 'A' levels Science Stream increased	Outcome	1	10	Government (SMSTU)
6	Enrolment in GCE 'A' levels Commerce Stream increased	Outcome	1	10	Government (SMSTU)
7	Principals and deputy principals trained	Output	2	32	Government (CELD)
8	Institutional capacity at Ministry of Education and provincial levels strengthened.	Output	3	32	Government (MOE/SMTSU)
9	Improved transparent and efficient procurement	Output	3	32	Government (SMSTU), ADB

Notes: CELD = Center for Education Leadership Development, DOE= Department of Examinations, GCE = General Certificate of Education, MOE = Ministry of Education, SMTSU = Sector Monitoring and Technical Support Unit.

Sources: Own compilation based on ADB (2013a)

20 The ADB uses the term 'disbursement-linked results' for clusters of results that are measured by 'disbursement-linked indicators'.

Focus on results

The results-focus of the programme can be assessed as relatively weak (see Table 6). Only 4 of the 21 DLIs measure outcomes and 20% (USD 40 million) of the loan is linked to the outcome-level indicators. In addition, 6½ indicators measure outputs, accounting for 40% of the loan. The remaining 40% of the loan (USD 80 million) are disbursed once certain input targets have been achieved or activities have been undertaken. The payment structure closely mirrors the results chain of the programme. Progress towards the nine intermediate objectives in the first two years is mainly measured by process indicators. In later years, an increasing number of results indicators, which are more direct measures of the nine results, are used. For instance, in 2013 and 2014, 70% of possible disbursements are linked to inputs and activities while in 2016 and 2017, 35% of disbursements are linked to output-level indicators and 50% to outcome-level indicators.

The predominant use of process indicators in early years of the programme allows a more gradual disbursement of funds and satisfies the need for pre-financing. It also shows the relatively strong engagement of the ADB in programme design and implementation. The necessary steps towards achieving the desired final outcomes of the programme are specified by the DLIs at input-, activity- and output-level and there is only little scope for the Sri Lankan government to experiment and to deviate from the prescribed path of development. It is particularly striking that DLIs under Objectives 8 and 9 (see Table 7), accounting for nearly one-third of total possible disbursements (USD 64 million), measure inputs, activities and outputs that are necessary to improve programme performance and to mitigate risks (the M&E system and procurement system) rather than results directly related to the overall objectives of the programme.

While the programme specifies the inputs and activities in detail, it aligns well with government priorities specified in the Sri Lankan government's Education Sector Development Framework and Programme (see Ministry of Education Sri Lanka, 2013).[21] The activities and outputs measured by DLIs are part of the government's development framework. The alignment of indicators shows that there is relatively strong agreement between the ADB and the government on how results can best be achieved. This justifies the extensive use of process level indicators to a certain extent.

Control

The majority of DLIs measure inputs, activities and outputs and are thus directly under control of the implementing agencies, namely the Ministry of Education and the nine provincial education departments. Only the eight DLIs at outcome level, measuring enrolment rates and pass rates, are subject to major external influences, thus making attribution difficult. In addition, improvements in outcome indicators might require several years before effects become observable, e.g. in the form of upgraded schools, additional teacher training in science, and exam pass rates. It can be assumed, however, that incentivised actors, given that they are the main government institutions responsible for education policies, are able to influence results at least to some extent. Targets for

21 Exceptions are indicators under Intermediate Objectives 8 and 9.

outcome-level indicators are low[22] and it can thus be expected that they are achievable over the programme timeframe with reasonable efforts.

The programme also includes provisions to establish a clear line of sight between the goals of the Ministry of Education and the provinces. DLIs under Intermediate Objective 8 measure whether performance-based partnerships that reflect alignment between the central and provincial annual work and budget plan were signed between the Ministry of Education and the provinces. These partnership agreements include necessary institutional arrangements to implement and monitor the programme, performance targets, respective responsibilities in delivering the results, and necessary financing as the incentives for provinces. Considering that provinces are responsible for the teachers and the majority of school resources, it is important for the Ministry of Education and the provinces to share responsibilities and have clear lines of accountability in programme implementation.

To further increase the probability that results are achieved within the programme time frame, DLIs measure whether at least 80% (years 2014 and 2015) or 85% (years 2015 and 2016) of the initially approved capital budget for school education has been spent. In the past, the Sri Lanka government, although fully committed to education, has suffered from a risk of low or delayed budget releases.

Financial incentives

When pricing results, the ADB has decided to link equal amounts of payments to individual indicators. Indicators that measure progress towards the five output-level objectives are each linked to USD 8 million during the first three years and USD 4 million during the last two years of the programme. All outcome-level indicators are worth USD 5 million of total funding. This decision helps to focus attention on the entire results framework and incentivises key reforms (leverage effect). Risk and 'value-for-money' considerations were not primary factors when deciding on the amount of funding linked to individual indicators. Generally, disbursements in ADB RBL programmes should not exceed total programme financing (ADB, 2013b, p. 28), so that value for money of the programme can be taken as given.

All disbursements in the ESDP are conditional on achieving a certain target or threshold value. Targets for process and outcome indicators vary in their level of ambition. The decision to not reflect the level of difficulty or value for money in the amount of financing linked to individual indicators may lead to a neglect of indicator targets that are not easily achievable. At the same time, it can be expected that the generally high levels of financing linked to each indicator create strong incentives for performance in all areas (leverage). The outcome-level targets are not very ambitious. While the majority of funding in the last two years is linked to outcomes, this only accounts for 20% in overall funding, contributing to a focus on activities and outputs rather than on the desired outcomes.

The programme stipulates that any amounts not disbursed for unmet DLIs will be disbursed once they have been achieved. This has various different implications: on the one hand, it weakens the incentive to achieve targets on time; on the other hand, there is incentive for

22 GCE 'O' and 'A' levels pass rates: 2% (4-5%) increase in pass rates until 2016 (2017); enrolment in Science and Commerce Stream: 4% (6-7%) increase until 2016 (2017).

improvements, even if achievement of the target seems rather difficult within the given time-frame. By specifying that at least three of the annual DLIs have to be met, it is moreover ensured that the recipient undertakes efforts to achieve at least some of the targets on time.

Measurability and verifiability

The two outcome-level objectives, increased enrolment rates and pass rates, are measured in a direct way by four quantitative outcome-level indicators. Progress towards the remaining five intermediate objectives at output level is measured by several input-, activity- or output-level indicators. Most are direct measures of the result (9½ indicators) and 7½ indicators can be classified as indirect indicators.

Data on DLIs measuring inputs, activities and outputs is provided by the ministries and institutions responsible for programme implementation and financing (i.e., the Ministry of Education; the Center for Education Leadership Development; and the Ministry of Finance and Planning). To measure development outcomes (pass rates for GCE 'O' and 'A' levels; enrolment in GCE 'A' levels Science and Commerce Stream) administrative data of the Sri Lankan Department of Examinations is used.

It can be expected that data on all indicators can be provided on a timely basis and that changes can easily be tracked. However, since several of the indicators are qualitative, there is some room for interpretation in the verification process. For example, one of the indicators measures whether time-bound development plans for the 250 selected schools have been completed and submitted to the Ministry of Education. As part of the verification process it has to be decided whether plans submitted include necessary details and fulfil relevant quality criteria.

While verification mechanisms and protocols have been established for all DLIs, only a few selected indicators under Intermediate Objectives 3 (Technology stream introduced and implemented at GCE 'A' Levels) and 4 (Secondary Schools upgraded to offer the Arts Stream, Commerce Stream, and Science Stream) are independently verified.[23] The remaining indicators are verified by government entities that are also responsible for programme implementation. For example, the majority of input, activity and output indicators are verified by a unit within the recipient's Ministry of Education, the Sector Monitoring and Technical Support Unit (SMTSU).[24] The establishment of the SMTSU is a DLI itself and part of the disbursement-linked result 'institutional capacity at MOE and provincial levels strengthened'. The SMTSU is not only responsible for overseeing the ADB programme but also for overseeing the implementation of the government's Education Sector Development Framework Plan (ADB, 2013b, p. 7). Similarly, the outcome-level indicators measuring enrolment and pass rates are verified by the Department of Examinations, which simultaneously is the source of data and responsible for managing examinations. For reporting from government entities, ADB's review missions work as spot checks for verifying data validity.

23 The ADB states that not all DLIs need independent verification. Where government has well-established systems for monitoring and capturing reporting, additional resources or efforts to independently verify might not be warranted.

24 Exceptions are the DLIs used to measure progress towards Intermediate Objective 7 (Principals and deputy principals trained). Here, data is monitored and verified by the Center for Education Leadership Development, which is also responsible for carrying out the training.

Using existing administrative data as well as relying on government entities for the verification of data is positive from the viewpoint of the aid effectiveness agenda. However, the relative lack of independent verification mechanisms questions the validity of results and implies a relatively high risk of data manipulation for both process and results indicators.

Unintended consequences

The strong focus of the ADB programme on inputs, activities and outputs combined with few easy-to-reach outcome-level targets minimises the risks of gaming and distortions.

The ADB decided to link a maximum of 20% of funding to outcomes (examination results), in order to minimise adverse incentives. The ADB argues that even though the examination results are controlled by an agency outside, the Ministry of Education, there is a risk of data manipulation because governance systems are vulnerable. The risk of data manipulation could have been reduced by the ADB by introducing systems for independent data verification. The risk that resources are reallocated from other potentially more important programmes to achieve ADB performance targets is comparably low, given the design of the programme. Funds are disbursed gradually, first (mainly) for activities and then increasingly for outputs. Outcome-targets only become relevant in the last two years. Payments for activities and outputs in the first three years are generous and first targets are achievable with a low budget. Hence, it can be expected that financing is not a major issue and resource distortions are not introduced.

At the same time, linking only 20% of funding to outcome indicators means that there is a low incentive to focus on the overall objective. There is the risk that the recipient concentrates on activities and outputs measured by DLIs and neglects other important actions needed to achieve outcome-level targets. The ADB aims to mitigate this effect by following up on a wider number of activities and outputs specified in the Programme Action Plan. This wider set of activities is important to achieve results, but would be too difficult to specify in the form of DLIs due to control or measurability issues.

4.1.3 DFID “Results Based Aid Pilot – Education Sector Programme” Rwanda

The results-based aid pilot in Rwanda is part of DFID’s Rwanda Education Sector Programme, which is embedded in the Government of Rwanda’s Education Sector Strategic Plan. The results-based aid component therefore only constitutes one element of DFID’s overall support to the education sector. Over the period 2011-2015, this support mainly consists of sector budget support in the form of an Education Service Delivery Grant (GBP 78.53 million), an Innovation and Capacity Building Fund (GBP 16.7 million) and a Technical Assistance Fund (GBP 1.85 million). Out of the Education Service Delivery Grant, up to GBP 9 million will be disbursed through the Results Compact (2013-2015) – the results based aid pilot.

This results compact aims to incentivise both improvements in student completion at key stages and improvements in teacher competency in English. It makes annual disbursements based on two main components: (1) additional numbers of students sitting exams (three indicators: primary, lower secondary and upper secondary school exams) and (2)

additional number of teachers reaching an agreed level of English competence (one indicator: English language test). Three disbursements of up to GBP 3 million were planned, and two disbursements of GBP 1.16 million in 2013 and GBP 1.88 million in 2014 were made. For these two disbursements, only the exam results were used as indicators. In 2015, the indicator assessing the English proficiency of teachers will be added. Table 8 shows an overview of the indicators and main features of the programme.

Table 8: DFID "Results Based Aid Pilot – Education Sector Programme" Rwanda

	Input	Activity	Output	Outcome	Total
Number of indicators	0	0	1	3	4
Amount of financing					
Incentivised actor	Ministry of Education				
Million GBP	0	0	0.155[a]	8.845	9
% of financing	0%	0%	2%	98%	100%

Indicator typology			
Individual	4	Quantitative	4
Composite	0	Qualitative	0

Notable characteristics				
Overall ceiling (million GBP)	9	Independently verified indicators	4	Independent Evaluation
Contract period (years)	3	Indicators verified by donor / government	0	

Outcome indicators

– Students sitting primary, lower secondary and upper secondary school exams

[a] This number is based on the indicator 'English language competency of teachers' which will be rewarded with a one-off payment in 2015. The amount here is calculated based on the assumption that 10% of teachers currently below envisioned level of English proficiency will improve towards the desired English level, which would represent a very substantial achievement (Health & Education Advice & Resource Team [HEART], 2013).

Sources: Own compilation based on DFID (2011c); HEART (2013); Musker, Clist, Abbott, Boyd, & Latimer (2014)

The programme in Rwanda features an independent evaluation. A first evaluation report has been published and a final evaluation report will follow in 2015. The first evaluation (Musker et al., 2014, p. 6) found that in Year 1 the RBA pilot *"did not make a significant contribution to the observed increase in completion rates in 2012"*, also because the RBA agreement was only signed in October 2012. The evaluation (Musker et al., 2014, p. 6) concluded it was *"too early in the implementation of the RBA pilot to expect results or assess value for money"*. Further, the evaluation report found that the RBA is characterised by a high-level of ownership and strong alignment with government priorities (Musker et al., 2014).

Focus on results

The results focus of the programme is relatively strong. The programme closely follows the Cash on Delivery model, and mostly rewards improvements in the number of students sitting exams. The three indicators measuring the number of students sitting exams are considered as outcome indicators here, although they do not consider exam passers. The advantage of using the indicator 'sitters' lies in the relatively easy verification and a relatively low potential for disputes (DFID, 2011b). A disadvantage could be that participation is increased regardless of the performance of students. The fourth indicator 'teacher's language proficiency' is classified as an output indicator because it represents a necessary step towards improving learning outcomes. This indicator was included in the programme design at the request of the Government of Rwanda (GoR), since there was a transition to English as the language of instruction in 2008.

Control

The DLIs measure outcomes over which the Ministry of Education has influence and reflect existing government priorities. The ministry has overall responsibility for budget execution and delivering education outcomes, but is also limited in several ways. For instance, the capacity of the Ministry has remained weak despite the capacity-building and technical support provided (DFID, 2011c). Further, management of the education system has been decentralised with many functions transferred to the districts, sectors, schools and communities. In financial terms, Rwanda has allocated between 15-17% of the government budget to education since 2007, which is lower than its own target value of 20% (Musker et al., 2014). Further, parents are key actors in the Rwandan education system and fund a significant share of education costs.

Within the government structures, the Ministry of Education should therefore be able to exercise influence on the number of students sitting exams. However, capacity or financial limits may negatively impact on the Ministry's ability to influence education outcomes. Also, the first impact evaluation notes that so far there is no evidence of new strategies or policies as a direct result of the RBA pilot. However, there might be external factors, other than the influence of the Ministry of Education, that influence the rate of exam sitters. For instance, the organisational culture in the education sector in Rwanda was already perceived as results-oriented prior to the introduction of the RBA programme. For instance, the existing system of so-called local *imihigo* performance contracts between districts and central government demonstrates this. Finally, the evaluation has shown that the RBA mechanism is little known at district and school levels, and only recognised at the central level.

Financial incentive

The baseline for a student sitting national examinations is linked to government administered tests in 2011 for Year Six in primary, Year Three and Year Six in secondary school exam sitters. Disbursements are made annually for improvements in the number of students participating in these three exams, with slightly different amounts:

- For each additional child sitting the P6 [Primary 6] exam above the previous year's results, DFID pays the GoR GBP 50. In addition, in years 2014 and 2015, DFID will

also pay GBP 10 for each additional child sitting the P6 examination above 2011 levels.

- For each additional child sitting the S3 [Secondary 3] exam above the previous year's results, DFID will pay the GoR GBP 100. In addition, in years 2014 and 2015, DFID will also pay GBP 10 for each additional child sitting the S3 examination above 2011 levels.

- For each additional child sitting the S6 [Secondary 6] exam above the previous year's results, DFID will pay the GoR GBP 50. In addition, in years 2014 and 2015, DFID will also pay GBP 10 for each additional child sitting the S6 examination above 2011 levels (Musker et al., 2014).

Further, the GoR receives GBP 50 for each additional teacher reaching an agreed level of competence by 2014 in comparison to a baseline in 2012. Disbursements will be calculated on this basis and are only made once in 2015, which can roughly be estimated to be around GBP 155,000. The major part of the total of GPB 9 million is allocated over three years for the outcome indicators, and there is an annual ceiling of GPB 3 million. While early DFID documents confirm the annual ceiling, later documents now tend to speak of GBP 9 million over three years. In either situation, naming a ceiling creates an implicit target for the partner to achieve.

The ambition of the programme to reward additional exam sitters by using two baselines, one for 2011 and then an updated dynamic baseline, is high. First, GBP 10 are paid annually in every grade for every additional student compared to the fixed 2011 baseline. Second, higher rewards of GBP 50 and GBP 100 are paid for additional year-to-year exam sitter based on the dynamic baseline. This incentive structure guarantees a modest incentive for continuous improvement (GBP 10) compared to the start of the programme in 2011, coupled with a more significant incentive for annual additional improvements (GBP 50; GBP 100). The allocation of rewards also shows that the highest value is placed on secondary third-year students, as the reward for each is GBP 100 compared to GBP 50 for primary year six and secondary year six sitters.

It is still open how unspent funds will be allocated at the end of the programme (currently GBP 5.95 million). According to programme provisions these funds could either be rolled over or be recycled into the innovation fund that goes to NGOs. This would not reduce the incentive effect of the results component towards the government, but would create implicit competition between NGO and government funding. Further, on the side of DFID, the RBA component (of up to GBP 9 million) only represents a small share of the overall Education Service Delivery Grant (GBP 78.53 million), which taken together with other forms of DFID support still represent less than 20% of the Rwandan education sector budget. The main constraint to achieving a stronger effect in terms of outcomes, however, might be the rather short time frame of three years of the RBA pilot.

Measurability and verifiability

The indicators for exam sitters are direct and quantitative indictors for school completion but indirect indicators for measuring learning outcomes. The indicator 'teacher's English proficiency' is a direct and quantitative indicator for measuring the learning outcomes of

teachers but a proxy for student learning outcomes. The outcome indicator ‘exam sitters’ was specifically picked to highlight school completion as this was considered to be the primary challenge in Rwanda. Indicators measuring the number of students passing key stage exams, which are more direct indicators of students’ learning outcomes, were not used because this might induce unintended consequences (see below). For the indicator assessing exam sitters, the administrative burden of financial management is considered to be minimal because these indicators rely on GoR systems.

The English proficiency of teachers is an output indicator that is a rather weak proxy for learning outcomes, but was introduced based on a request of the GoR and only represents about 2% of the programme budget. The English proficiency of teachers is assessed by a test administered by the British Council Rwanda according to the Common European Framework of Reference (CEFR) for Languages levels. The assessment of English proficiency is therefore not based on existing systems in Rwanda but has been introduced specifically for the RBA pilot.

Verification is carried out by a consultancy association called Health & Education Advice & Resource Team (HEART), which is a consortium of organisations in international development. The independent verification has two dimensions: administrative and procedural (HEART, 2013). The team conducting the verification is independent of both DFID and the GoR administratively, thereby ensuring accountability, transparency and potentially greater confidence in the findings. Also, the verification procedures involve cross-referencing the results to be verified with independent evidence in this case, schools’ records relating to those results. The verification team mostly relies on data generated by Rwandan authorities but cross-checks this data by conducting field surveys. For the baseline study, the verification team concluded that the 2011 examinations statistics provided by the Rwandan authorities were accurate and reliable, as less than 1% of national data and school-level data showed discrepancies (HEART, 2013).

Unintended consequences

For the RBA pilot programme in Rwanda, the programme documents at the design stage explicitly cite unintended consequences (such as distortion, cherry-picking or gaming) as potential challenges in implementing RBA. As a mitigation measure against gaming, DFID has set up an independent verification process. In addition, the evaluation of the programme was tasked to investigate unintended consequences.

The evaluation report mentions potential perverse incentives generated by the RBA pilot. For instance, the DLIs measuring the number of pupils sitting key stage exams are useful to measure completion of education but not to measure learning outcomes. Thus, a strong focus may be put by the government on increasing education access and retaining students in school but not on increasing the quality of education. In addition, the government might push for better completion rates by pressuring school principals to increase the number of sitters, regardless of whether students are actually enrolled. However, the evaluation report notes that GoR *“has responded to the RBA incentive in a manner that counteracts possible perverse incentives by insisting that students enroll and attend school in order to qualify for sitting the key stage examinations”* (Musker et al., 2014, p.40). In addition, the evaluation states that *“there is no evidence that the RBA approach has resulted in*

practical changes in policy implementation 'on the ground' that result from perverse incentives and detract from sector-wide government plans" (Musker et al., 2014, p. 43).

4.1.4 DFID "Pilot Project of Results Based Aid in the Education Sector" Ethiopia

The DFID "Pilot Project Results Based Aid in the Education Sector" in Ethiopia aims to improve access to and quality of lower secondary schooling. The project is implemented by the Ministry of Education, which is also the recipient of payments. Funding provided by DFID as part of the project is additional to DFID's education sector support[25] and does not substitute for input-based funding (DFID, 2011b). In total, GBP 30 million are available for disbursements over the project period (2011/2012-2014/2015), with an annual ceiling of GBP 10 million. The programme is aligned with the government's development plan and addresses one of the government's main priorities.

The expected theory of change of the project is as follows. The intended outcome of the project is an increase in Grade 10 students sitting and passing the Grade 10 examination, especially for girls and in emerging regions. The expected outputs are:

(a) The government adopts policies and programmes that lead to increased enrolment and retention of students in lower secondary school

(b) The government uses existing resources in a more efficient and targeted way.

The project relies on the assumption that incentives of the RBA programme will trigger a policy response not only at federal level but also at regional, *woreda* (district) and school level. For this to happen, the Ministry of Education has to create incentives for other levels of government to contribute to the project's objectives. As part of the Memorandum of Understanding, it was agreed that the Ministry of Education is in charge of developing guidelines for how money will be allocated to regions and for how it should be used by regions. Regions should in turn decide on how to allocate additional funding further to districts and schools. Only if the details of the RBA are communicated to regions, and if there is clarity about the allocation of payments, will incentives of the RBA trickle down to local levels of government.

Table 9 shows an overview of the types of indicators used and the main characteristics of the programme.

25 DFID provided GBP 209 million education sector support to Ethiopia over three years from 2011/2012 – 2013/2014 (DFID, 2014b, p. 11).

Table 9: DFID "Pilot Project of Results Based Aid in the Education Sector" Ethiopia

	Input	Activity	Output	Outcome	Total
Number of indicators	0	0	0	2	2
Amount of financing					
Incentivised actor	Ministry of Education				
Million GBP	0	0	0	30	305
% of financing	0%	0%	0%	100%	100

Indicator typology			
Individual	2	Quantitative	2
Composite	0	Qualitative	0

Notable characteristics				
Annual ceiling (million GBP)	10	Independently verified indicators	2	Independent Evaluation
Contract period (years)	3	Indicators verified by donor / government	0	

Outcome indicators
– Additional students sitting the national Grade 10 examination – Additional students passing the national Grade 10 examination

Sources: Own compilation based on DFID (2011b)

Focus on results

The results focus of the programme can be assessed as strong. The programme, in line with the Cash on Delivery Aid model[26], rewards additional, incremental changes and uses two DLIs, both at outcome level, which are direct and quantitative measures of the desired results:

- Students sitting the national Grade 10 exam
- Students passing the national Grade 10 exam

In line with the COD Aid approach, the project does not use process-level DLIs to ensure that certain pre-defined measures are undertaken by the Ministry of Education to achieve

26 There are some important differences between the COD Aid approach and the DFID pilot on RBA in Ethiopia: in a 'pure' COD Aid approach, there would be no annual ceiling for disbursements, funding would be unearmarked and go to the National Treasury or Ministry of Finance (not to the Ministry of Education), there would only be one national baseline to calculate additional achievements, and only one unit of reward (no disaggregation between groups and regions). Finally, test scores would not be used to disburse aid because of concerns about the volatility of country learning assessments over time (Birdsall & Perakis, 2012, p. 3; DFID, 2011b).

the results. In contrast, it is left to the Ministry of Education to decide which path to follow to achieve the agreed results. Thus, the pilot was designed to allow for a high level of government ownership. First evidence from the DFID pilot in Ethiopia, however, shows that the government was slow to take the initiative and that the policy response was limited. For example, according to DFID (2013b, p. 6), it is unclear whether the government would have undertaken some of the activities seen as essential by DFID to achieve results, such as communicating to regions about the RBA design and developing guidelines for how to distribute funds to regions, without the initiative and capacity development support provided by DFID.

Control

The DLIs measure outcomes over which the Ministry of Education has a reasonable degree of influence.[27] The factors influencing the results are, however, numerous. The policy response is expected at different levels of government. Hence, whether or not progress can be made depends to a large degree on the willingness and capacity of regions, districts and schools to contribute. The low time horizon of only three years limits the possibility of making changes. It can be expected that reforms to improve completion rates and learning outcomes take several years to bear fruit.

In addition, financial constraints limit the capability of the Ethiopian government to achieve results. The Ethiopian government estimates the funding gap in the National Education Sector Development Plan IV to be USD 1.1 billion which is expected to be only partly filled by donor contributions[28] (Federal Democratic Republic of Ethiopia, 2010, p. 102). Besides, there is only little room to increase domestic funding for education. The government already spends 5% of its gross domestic product and 20% of its federal budget on education (DFID, 2011b). As the need for pre-financing is not addressed by the RBA, it is questionable whether the government will be able to make the necessary investments to increase the number of students sitting and passing the national Grade 10 exam. Acknowledging the financial constraints of the government, DFID's theory of change relies on the assumption that existing resources will be used more efficiently (DFID, 2011b). This alone may, however, not be sufficient to achieve results.

Financial incentives

DFID pays for each student who sits/passes the Grade 10 examination over an agreed baseline. In addition, a premium is paid for students in emerging regions[29] and for girls

27 Before the start of the programme, DFID evaluated the capability of the Ethiopian government to increase the number of students sitting and passing the national Grade 10 examination as relatively high, given Ethiopia's strong track record in the education sector. Ethiopia was able to increase enrolment in primary education by 2 million and enrolment in lower secondary education by 500,000 over the previous five years (DFID, 2011b).

28 Donor contributions are estimated at approximately USD 310 million (Federal Democratic Republic of Ethiopia, 2010, p. 102).

29 Ethiopia's four developing regional states, i.e. Somali, Afar, Gambella, and Benishangul Gomuz, are classified as emerging regions. These are underdeveloped compared with other regions, have a relatively high share of pastoral communities and in some cases are prone to conflict (DFID, 2011b).

(see Table 10). Baselines have been calculated separately for each of the eleven regions[30] and for girls and boys.

DFID took two considerations into account when deciding on the amount of payment: value for money and leverage. The payment for boys in non-emerging regions is based on the actual costs for educating a secondary student for the two years in lower secondary education. The premiums for girls and students in emerging regions are paid to incentivise the government to initiate and implement programmes that focus on the latter in particular (DFID, 2011b). Both groups are less likely to be enrolled in secondary schools and at the same time less likely to pass the national Grade 10 exam (DFID, 2011b). [31]

Table 10: Payment structure

In emerging regions DFID pays	**In non-emerging regions DFID pays**
… for each additional student sitting the national Grade 10 exam (boys: GBP 75, girls GBP 100)	… for each additional student sitting the national Grade 10 exam (boys: GBP 50, girls GBP 85)
plus an additional … for each of these students who pass the national Grade 10 exam (boys: GBP 75, girls GBP 100)	plus an additional … for each of these students who pass national Grade 10 the exam (boys: GBP 50, girls GBP 85)
Source: DFID (2011d)	

The annual reviews of the project carried out by DFID in 2012 and 2013 question whether financial incentives provided by the programme have had the intended effect. The government focused more on deciding how additional funds could be allocated than on developing and implementing policies that increase the number of students sitting and passing the exam. There is no evidence yet that the RBA programme has led to a greater efficiency in the use of existing resources (DFID, 2013b, 2013c). Results are rather disappointing. The total number of students sitting and passing the Grade 10 examination even fell from 2011 to 2012 (DFID, 2013b). In 2013, the number of sitters increased compared to 2012 but still remained below the 2011 level. The number of passers was above 2011 and 2012 levels in 2013 (DFID, 2013b, p. 6; 2013c, p. 16).[32]

Still, there is some indication that the programme has had the intended effect of increasing the focus on emerging regions and girls. In 2012, four of the six regions that were able to increase either sitters or passers among either boys or girls were emerging regions. In addition, 74% of the additional sitters and 99% of the additional passers were in emerging

30 Addis Ababa, Afar, Amhara, Benshangul Gumuz, Dire Dawa, Gambella, Harari, Oromiya, Somali, the Southern Nations, Nationalities, and Peoples' Region (SNNPR), Tigray

31 In 2008/2009, the net enrolment ration for students in emerging regions was very low at 2.7%. Only 34% of students from emerging regions passed the Grade 10 examination in that year, compared to the national average of 43%. Girls accounted for 43% of students entering grade nine in 2008/2009 but only for 37% of students passing the national Grade 10 examination in the subsequent year (DFID, 2011b).

32 DFID pays for improvements over the previous year. If the 2011 performance level was used as a baseline during the entire project period, as recommended by the Center for Global Development, payments made by DFID in 2013 would have been much lower because performance was reduced from 2012 to 2011. This is the opposite of what DFID expected. DFID chose an adjusting baseline because of concerns that using only 2011 levels as a baseline could result in financial liabilities beyond the funds available if performance increased from year to year (DFID, 2013b, p. 17).

regions (DFID, 2013b, p. 6). In 2013, improvements in sitters and passers over the 2011 baseline were seen for girls, but not for boys (DFID, 2013c, p. 16). Still, it is difficult to say whether these changes are an effect of the RBA pilot or would have occurred anyway even in the absence of the programme.

One question that arises is whether potential payments are large enough to generate incentives for the government. The GBP 10 million that DFID allocated to the project per year are low compared to total annual donor contributions to the government's Education Sector Development Plan of approximately USD 300 million (Birdsall & Perakis, 2012, p. 3; Federal Democratic Republic of Ethiopia, 2010, p. 102). The annual ceiling of GBP 10 million also acts as an implicit target for the government. When only GBP 896,260 were disbursed by DFID in the first year, it may have discouraged the government to further invest in achieving the RBA results due to the low returns compared to what was expected.

Measurability and verifiability

The indicators 'number of students sitting' and 'number of students passing' the national Grade 10 examination are both direct and quantitative measures of the intended project outcome.

For both indicators, national data from the Education Management Information System is used. Grade 10 examinations are set and graded at the federal level and the results of the examinations are published annually in the Annual Statistical Abstract (DFID, 2011b). To mitigate the risk of cheating and to ensure data quality, data on both indicators, i.e. baseline data and annual results, is independently verified based on a random sample of secondary schools.

While using the indicator 'number of students sitting the national grade 10 examination' is generally seen as unproblematic, concern was expressed by DFID and the Center for Global Development about data quality and validity of the indicator 'number of students passing the exam'. One of the main challenges lies in ensuring that the Grade 10 examination is 'equivalent' – and hence comparable – in terms of difficulty from year to year (Birdsall & Perakis, 2012, p. 5; DFID, 2011a). In recent years, the level of difficulty of the Grade 10 examination has fluctuated. To ensure comparability, a review of the cut-off point for passing the examination is part of the independent verification (DFID, 2011a).

Unintended effects

The choice of the indicators in the DFID RBA pilot in Ethiopia may lead to several unintended effects.

Distortions may arise as a result of the emphasis on secondary education and the budgetary constraints of the government. Given the lack of pre-financing, there is a risk that the emphasis on secondary education diverts resources away from primary education or from other sectors. It could happen that the government does not undertake additional efforts and that DFID pays for results that would also have been achieved in the absence of the project.

Cherry-picking is another likely unintended effect of the project. DFID makes higher payments for girls and students in emerging regions, thereby lowering the risk that the government focuses only on those easiest to reach. At the same time, especially in the four emerging regions where results are difficult to attain and rewarded by high payments, the government may focus on easier-to-reach communities and neglect pastoralist communities in remote areas.

Gaming is another potential unintended consequence of the project. Achieving desired changes in outcome indicators within the short project period of three years is difficult because most efforts to build institutions or capacities will only have an effect over the long term. The government thus has an incentive to engage in activities such as data misrepresentation to ensure that payments are made.

To mitigate these kinds of effects, DFID carefully monitors the response of the Ethiopian government through an accompanying evaluation and the secondment of a senior education advisor to the Ministry of Education. This puts pressure on the Ethiopian government to focus on policies that contribute to achieving the agreed results.

4.1.5 Twaweza "KiuFunza Local Cash on Delivery" Tanzania

The overall aim of the intervention "KiuFunza" ('Thirst to learn') is to improve early grade learning outcomes (literacy and numeracy) in Tanzania. KiuFunza is a 'local Cash on Delivery' pay for performance intervention that builds on the COD Aid approach. In contrast to COD Aid, KiuFunza is not a contract between governments. The intervention is managed by a local NGO (Twaweza), supported by several partners[33], and payments for results are made directly to teachers. It is carried out on a pilot basis, tested in 14 schools in 11 districts only, and organised as a randomised control trial (RCT) to test the effectiveness of the approach (COSTECH, EDI, JPAL-MIT, & Twaweza, 2012, pp. 2-3).[34]

Since KiuFunza is a pilot project, the amount budgeted for disbursement, USD 195 thousand annually, is low. According to estimates, the intervention would cost USD 45 million per year, if implemented country-wide and if every child were to pass (COSTECH et al., 2012, p. 5). The design of the COD intervention has been kept simple: only three indicators are used, incentives and responsibilities are clear, there is a strong focus on improving learning outcomes in the three tested subject areas, and results can easily be communicated to the public. The simple design allows for the easy engagement of

33 Several partners support KiuFunza. Among them are the Government of Tanzania, the Prime Minister's Office (Regional Administration and Local Government), the Ministry of Education and Vocational Training, the Ministry of Finance and Economic Affairs, COSTECH, the Tanzanian Teachers' Union, the Center for Global Development (CGD), Abdul Latif Jameel Poverty Action Lab (J-PAL)/Innovations for Poverty Action (IPA), Uwezo, the Economic Development Initiatives (EDI) and several donors (including DFID, CIDA, SIDA, the World Bank (Costech et al., 2012, pp. 7-8).

34 Twaweza implements three experimental interventions to compare their effects: (1) the direct transfer of payments for inputs to school accounts (capitation grants), (2) the payment of a bonus to teachers for every child who passes a literacy (English, Kiswahili) and numeracy (mathematics) test at the end of the school year and (3) a combination intervention that provides the capitation grant and pays teachers for performance. The RCT is carried out in 11 randomly selected districts and in 21 randomly selected schools (Costech et al. 2012, pp. 1-2).

teachers, pupils, parents and also of policymakers in the intervention (COSTECH et al., 2012, p. 4).[35] Table 11 gives a brief overview of the programme and the indicators used.

Table 11: Twaweza "KiuFunza Local Cash on Delivery" Tanzania

	Input	Activity	Output	Outcome	Total
Number of indicators	0	0	0	3	3
Amount of financing					
Incentivised actors	Teacher and head teacher				
Million USD	0	0	0	0.59[a]	100
% of financing	0%	0%	0%	100%	100%

Indicator typology			
Direct	3	Quantitative	3
Indirect	0	Qualitative	0

Notable characteristics			
Results-based finance (Twaweza – teachers)	Independently verified indicators	3	Randomised control trial
	Contract period (years)	3	

Outcome indicators

- No. of children who pass a literacy (English) assessment at the end of the school year
- No. of children who pass a literacy (Kiswahili) assessment at the end of the school year
- No. of children who pass a numeracy (mathematics) assessment at the end of the school year

[a]Amount of financing calculated based on expected upper-bound pass rate: English 5%; Kiswahili 60%; maths 60% (50,000 students in COD pilot schools) (Twaweza, 2014). USD 195,000 are budgeted by Twaweza for 2014 disbursements. Because data for 2015 and 2016 are not available, we assume that the amount stays the same for both years.

Sources: Own compilation based on Costech et al. (2012)

Focus on results

The project focuses strongly on outcomes. Three DLIs at outcome-level are used (see Table 11), which are direct and quantitative measures of the desired result (improved learning outcomes). A payment is made to the child's teacher and the head teacher for every test a child in Grades 1, 2 and 3 passes in literacy (English, Kiswahili) and numeracy (mathematics) assessments at the end of the school year (Costech et al., 2012, p. 4).

Control

Teacher quality has been identified as one of the main factors influencing student achievements besides factors such as a student's cognitive ability and background, class size and peer learning achievements (Darling-Hammond, 2000; Nye, Konstantopoulos, & Hedges, 2004; Rockoff, 2004). It can thus be expected that additional efforts by teachers

35 Another reason provided by Costech et al. (2012, pp. 4-5) of why the design has been kept simple is that it could also be managed by a weak bureaucracy if scaled up nationwide.

lead to measurable changes in learning outcomes. Teachers can respond to the incentive by e.g. introducing new teaching methods, preparing lessons better, paying special attention to weaker students, and additional voluntary teaching activities. In Tanzania, where teacher classroom absenteeism is high[36] and teacher motivation low, one of the main responses may be increased attendance and higher teaching quality. Head teachers, who also receive a bonus for each child passing a test, have less control over children's learning outcomes. However, they can influence the behaviour of teaching staff. For example, they can take action to reduce teacher absenteeism, to increase the time teachers spend in class, and to improve teaching quality.

Financial incentives

KiuFunza makes payments directly to teachers based on the assumption that incentives are most effective when paid to those most at stake, and when potential payments are significant (leverage effect) (COSTECH et al., 2012, p. 3). A teacher receives Tanzania shillings (TZS) 5,000 (USD 2.8) for each test a child in his or her class (Grades 1, 2 and 3) passes (up to TZS 15,000 for a child who passes all three tests). In addition, the school's head teacher is paid TZS 1,000 (USD 0.56) for each child passing a test. For a teacher in Tanzania, payments may amount to a substantial bonus. A teacher teaching all three subjects in a class with 60 children could, for instance, earn up to USD 504, which equals almost two monthly salaries.[37]

When deciding on the amount of payments, Twaweza also took into account value-for-money considerations and the affordability of the intervention for the government if scaled-up nationwide (COSTECH et al., 2012, p. 5). The pass rate in learning assessments is currently at 20% (COSTECH et al., 2012, p. 5) and the intervention could still be afforded by the government if the pass rate increased to 60%.

To ensure simplicity of the intervention, Twaweza decided to pay for absolute achievements and not for gains in learning, i.e. teachers receive a payment for every test a child passes and not only for improvements above a baseline. Hence, teachers may receive payments, even if they do not undertake additional efforts for those pupils who would have passed anyway (COSTECH et al., 2012, p. 5). Paying only for the number of additional students passing a test over a baseline would have meant a greater incentive for teachers to improve over current conditions. Twaweza currently seeks to implement this aspect in a new design in 2015 and 2016.

Measurability and verifiability

The indicators chosen are direct and quantitative measures of the objective of improving early grade learning outcomes (literacy and numeracy).

---

36 The Service Delivery Indicators Report for Tanzania (World Bank and African Economic Research Consortium, 2012) shows that in 2012 about a quarter of teachers in Tanzania were absent from school on any given day and that even if teachers are in school they are absent from the classroom more than half of the time. This leads to low overall teaching times. Students are on average taught about 2 hours and 4 minutes per day.

37 Teachers earn TZS 575,000 (USD 306) per month, pre-tax, including allowances (Manda & Schipper, 2014).

Various different measures have been taken by Twaweza in response to concerns over the validity and reliability of national learning assessments and potential unintended effects. Tests administered to students are based on the Uwezo methodology. Uwezo is a five-year initiative under Twaweza that assesses literacy and numeracy levels for children aged 6-16 years by means of country-wide household surveys (Uwezo, 2014).[38] Uwezo tests are standardised and have been shown to be reliable and comparable in terms of difficulty between years (Jonesa, Schipper, Rutod, & Rajanib, 2014).

To ensure test integrity and to minimise the risk of distortions of a high-stakes test, several precautions have been taken by Twaweza. For example, ten different sets of tests were developed for each grade and subject and tests were randomly assigned to each student. Tests were conducted on a one-to-one basis by independent implementers without the involvement of teachers. Tests were all retained in the Twaweza office and results were analysed to identify manipulation at the tester, teacher and coordinator level (Manda & Schipper, 2014). [39]

Unintended consequences

The direct payment of performance incentives to teachers creates different risks. When teachers are paid for their students' performance on tests, tests become 'high-stakes' tests[40] because teachers have a high interest that students perform well (Lockheed, 2008, p. 13). Besides leading to improved learning outcomes, this may have various different adverse effects such as instruction that emphasises tasks that resemble tests, the neglect of weaker and hard-to-teach students during lessons, and ultimately inflated test scores without actual improvements in learning (Lockheed, 2008, p. 13; Shepard, 1991). To date, it was observed by Twaweza that the subjects tested, English, Kiswahili and maths, became more popular among teachers. This is an intended effect of the programme because Twaweza considers English, Kiswahili and maths as core skills that are necessary to succeed in other parts of the curriculum. However, the focus on the three subjects could still be considered as negative to some extent because it could lead to other important subjects being neglected.

4.2 Comparison of indicators used in the five selected programmes

The five pilot programmes provide a cross-section of how development organisations are implementing results-based approaches at present. Our analysis has presented an overview of the indicators used in each programme and has assessed the quality of DLIs according to several criteria. In this section we turn to cross-cutting findings that emerge from our analysis. We group these findings along the main criteria developed above and focus on trade-offs and challenges within each criterion.

38 To date three assessments have been carried out in 2009/10, 2011 and 2012 (Uwezo 2014).

39 It was found that manipulation occurred in two schools in one district at the coordinator level, where final results corrected for this manipulation (Manda & Schipper, 2014).

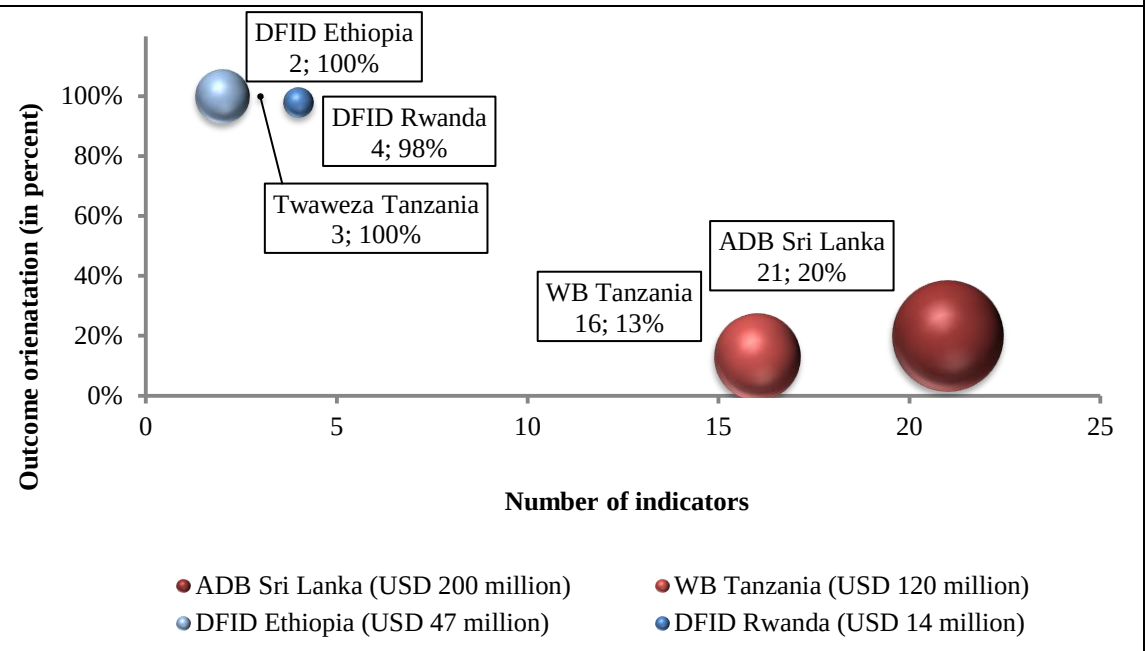
40 In comparison, 'low-stakes' tests are not used to make decisions about individuals or groups although they can be used to make a comparison between students to inform decision-making (Lockheed 2008, 13).

4.2.1 Focus on results

The five programmes are characterised by different degrees of results orientation (see Figure 2). The development banks offer a high amount of potential financing while taking a cautious approach by predominantly relying on process indicators. Only a small share of total funding (20% in the ADB programme and 13% in the World Bank programme) is linked to outcomes. In comparison, the three Cash on Delivery (COD)-type approaches provide relatively low amounts of financing and rely almost exclusively on outcome indicators. The DFID programme in Ethiopia and the Twaweza programme exclusively pay for outcomes, while the DFID programme in Rwanda in addition links a small share of funding to outputs.

The variation in outcome orientation has different implications. The approach of the development banks has the advantage that financing for implementing the programme can be provided early within the programme. A gradual approach is taken that links first disbursements to preparatory activities, such as setting up staff and procedures for implementing the results-based approach and to activities that are expected to contribute to the intended results. In later years, disbursements are made increasingly for outputs and outcomes. By relying on process indicators, however, the banks are strongly involved in the implementation process and leave the partner-side little room to innovate. Also, the higher the number of process indicators used, the more traditional and less innovative these programmes look. The outcome indicators used by the banks are formulated cautiously and only trigger relatively small disbursements at later stages of implementation. The modest outcome orientation makes sense when programmes follow a gradual approach of phasing in, but weakens the incentive effect of the outcome indicators overall programme.

Figure 2: Outcome orientation, numbers of indicators and programme budgets



Source: Authors

By focusing on outcome indicators, the three COD-type approaches leave a large degree of discretion to the incentivised counterparts in determining what means to adopt for achieving the RBA results. In both DFID pilots, Ethiopia and Rwanda, the RBA programmes are rolled out nationwide with set unit payments for increases in the number of exam sitters and passers. The Twaweza programme in Tanzania is implemented on a pilot basis in selected schools and pays teachers for improvements in student learning outcomes. While the use of outcome-level indicators leaves room for experimentation and theoretically allows for a high level of ownership, first evidence from the two DFID pilot programmes shows that new policies or strategies have not yet been introduced as a result of the RBA (Musker et al., 2014; Perrin, 2013).

Overall, programmes face a basic trade-off between a simple, outcome-focused design and a more complex process-focused design. RBA programmes that rely on outcome indicators are more innovative and leave more room for experimentation. The more results-based approaches focus on process indicators, the more they resemble traditional input-based aid programmes.

4.2.2 Control

Across the five programmes, the incentivised counterparts have different levels of control over achieving the intended results.

The strong focus on process indicators in the programmes of the development banks implies a high level of control of the recipient governments over achieving agreed results. The majority of indicators prescribe certain policy actions and outputs that have to be delivered in a specified manner and name the agencies responsible for implementation. Attribution of activities and outputs to the programme is possible. For the outcome indicators used attribution is not possible, but these indicators only represent a small share of total funding. Further, they are only relevant during later stages of implementation and it can be assumed that the relative lack of control is not a strong concern for the recipients.

The RBA programmes implemented by DFID on the other hand, mostly rely on indicators that measure outcomes in the form of exam sitters or passers. These indicators are subject to several influencing factors that cannot be controlled by the incentivised ministries. For instance, the number of students passing exams might fluctuate naturally due to economic, social and political factors outside the control of the RBA programmes. For the programmes to work, several conditions have to be met. The responsible ministries of education must have the capacity to initiate policies that are implemented at regional, local and school levels. A clear line of sight approach must be followed that links the objectives at the different levels of government to the objectives of the RBA programme. Yet, the ability of the education ministries to influence the results is limited by the short programme periods (3 years) and by government budget constraints (pre-financing is not provided).

The Twaweza pilot differs from the other programmes because it is implemented by a NGO, and disbursements are made directly to teachers. Although payments are also linked to outcome indicators (number of tests passed by students in learning assessments), the level of control is higher compared to the two DFID programmes. While teachers also rely

on the actions of others (students, parents, schools, education institutions) to achieve changes in learning outcomes, they have a more direct influence over student behaviour. Therefore, it might be easier to assess whether the RBA programme has contributed to significant changes in learning outcomes. Here, first evaluations have yet to be published, but it will be interesting to see the insights of the randomised controlled trial on the influence of teachers on learning outcomes.

Again, there is a trade-off between a simpler more outcome-focused design and a more complex, process-focused design. In outcome-focused programmes, there is a strong results orientation but in most cases, the ability to influence results is fairly limited, especially if programme periods are short. In process-focused programmes the short-term changes in policy processes can be more easily attributed to the RBA intervention. At the same time, there is only a weak link between the disbursement-relevant activities (leading to short-term results) and intended development outcomes.

4.2.3 Financial incentives

Different criteria have been taken into account by the programmes to decide on the amounts of payments linked to each DLI.

The DFID programmes aim to both maximise value for money and at the same time to achieve a leverage effect. For example, the unit prices in the DFID programme in Ethiopia are based on the actual costs of educating students for one year and a bonus is paid for girls and for students in disadvantaged regions who are more difficult to reach. By choosing only a few outcome-level indicators, DFID aimed to focus the attention of the government and to incentivise key reforms (leverage). Twaweza ensured that potential payouts for teachers are significant while taking into account government budget limits. For the development banks, leverage was the most important criterion. The total amount of financing is much higher than the amounts in the three COD-Aid programmes. However, value-for-money in the programmes of the development banks is still implicitly given because total disbursements in World Bank and ADB programmes should only cover a share of total programme expenditures.

The programmes of the development banks can be said to have leverage which can be explained by the high amount of total funding available. The Twaweza programme also creates strong incentives for improved performance. It was found that average payouts for teachers in year one were significant although pass rates were relatively low.

A second difference between the development banks, DFID and Twaweza lies in making disbursements according to targets or for improvements on a scale. While DFID and Twaweza solely use unit prices for improvements, the World Bank and ADB also set targets for indicators. By setting targets, the development banks create expectations for minimum achievements and create ‘all-or-nothing’ situations with downsides for all stakeholders. Even if partner countries miss the target by a small degree, they will not receive disbursements.

Financial incentives depend on how undisbursed funds linked to individual indicators are reallocated. The five programmes take different approaches to reallocating unspent funds.

The ADB allows for shifting funds from unmet DLIs to DLIs that have been achieved, thereby potentially undermining the results incentives. The World Bank has provisions for reallocating unspent funds to later years and shifting between DLIs, which could undermine the incentive scheme. DFID allows reallocation, but only to other DFID-financed activities in the education sector that do not benefit the government but non-governmental actors. Such a design is less likely to undermine the financial incentives set by the RBA programme.

With regard to financial incentives, a first trade-off can be identified between value for money considerations and maximising leverage by linking high payments to indicators. If value-for-money considerations lead to a large number of indicators that trigger relatively low disbursements, this would limit leverage. If larger amounts of funding are focused on a small number of indicators this could maximise leverage, but might not represent good value for money. A second trade-off exists between setting targets for indicators and disbursing for unit improvements. Targets, if well chosen, can maximise incentives for a strong one-time improvement but also create dangerous ‘all-or-nothing’ situations. Unit price disbursements reward a continuous level of improvement, but might induce only a modest level of ambition, especially in later years, when improvements are more difficult to achieve at the margin.

4.2.4 Measurability and verifiability

All five programmes aim at improving learning outcomes but feature different ways of measuring learning outcomes and verifying results. The development banks adopt a variety of DLIs. Only a few indicators are direct measures of the intended outcomes and several indirect indicators at input, activity and output levels are used, where the relation to underlying results is often quite weak. DFID and Twaweza only use direct indicators that are closely linked to the desired result at outcome-level.

For outcome indicators, measuring enrolment, pass rates and test scores, all programmes except for Twaweza use national administrative data. This can be seen as positive when judging from an aid effectiveness perspective, although in some cases the validity of results might be questionable. In particular, the use of examination results to disburse aid is not always advisable because tests scores tend to fluctuate widely between years, due a lack of standardisation. Different approaches have been chosen to address this problem. For example, the DFID programme in Rwanda has decided against the use of examination results to disburse funds and instead uses a DLI measuring the number of students sitting exams. This indicator is a weaker proxy for student-learning outcomes but tends to be more reliable. Twaweza does not use administrative data to measure learning outcomes, but tests students using a new system and methodology that has been shown to produce reliable results. The development banks have only linked a small share of total funding to outcome indicators which reduces the risk of unreliable measures for both the banks and the partner governments. The World Bank chooses to adopt 3R assessments in Tanzania as an outcome indicator because it is part of a newly introduced reform agenda (Big Results Now) and promises to provide more stable and reliable results than existing student assessments.

Different methods are applied to verify data by the programmes. In the three programmes based on the COD Aid approach, data is independently verified. In contrast, the development banks manage parts of the verification process by themselves, some indicators are verified by the partner governments and only selected indicators are independently verified. Arguments against independent verification might be cost aspects or a strong interest of either funder or the partner country to control the verification process. However, experience with results-based approaches shows that independent verification is integral for ensuring a strong incentive effect and credibility of the programme. Independent verification is also needed to ensure confidence in findings and to improve accountability and transparency.

A central trade-off related to measurability is the choice between relying on country systems and administrative data or introducing new measurement systems that either create or duplicate existing structures. Setting up new structures goes against the ownership and alignment principles of aid effectiveness, but might guarantee more reliable data. Using a country system conforms to the aid effectiveness principles but might yield unreliable data.

4.2.5 Unintended consequences

The risk of unintended effects and distortions through gaming applies to all of the five programmes studied. Different unintended effects are, however, more likely to occur in some programmes than in others, depending on the indicators used and the type of programme.

The risk that the recipient engages in activities that contribute to increasing the value of a DLI but not the overall objective (measure fixation) is generally higher in the three COD-type interventions. The three COD-type programmes leave full discretion to the partner on how to achieve the result, whereas the development banks specify the activities and outputs needed to achieve a desired result. In the latter case, the partner has a strong interest in undertaking activities prescribed by the programme and there is only a low incentive to engage in additional activities that may not be beneficial to the overall objective. On the other hand, the focus on process indicators in the programmes of the development banks implies the risk that long-term goals may be neglected.

If imperfect performance indicators are used to measure outcomes, the risk of measure fixation is particularly high. The actions required for increasing the value of the indicator might be different from actions that are needed to achieve the underlying programme objective. The DFID pilot programme in Rwanda might especially be prone to this risk. The programme aims to improve the quality of education, but only uses DLIs that measure the number of students sitting key stage examinations, which is a measure of school completion rather than quality of education.

The risk that resources are diverted from other sectors or programmes to the results-based approach is higher in COD Aid-type interventions. Unlike in the programmes of the development banks, payments are only made after improvements in outcomes have been achieved and pre-financing is not provided. If a more efficient use of resources is not

possible, reallocation of resources and a shift in priorities may be necessary to achieve agreed results.

Generally, the short contract periods in the COD-type programmes[41] increase the risk of adverse effects. Achieving desired changes in outcome indicators within a few years is difficult because most policy changes or programmes to build institutions or capacities will only have an effect over the long term. If there is a low probability that results are achievable with genuine efforts within the contract period, then the incentive for gaming is high.[42]

The risk of data manipulation can be assessed as higher in the programmes of the development banks. While all indicators in the COD-type interventions are independently verified, this is only done for a few indicators in the programmes of development banks. Most indicators in both the World Bank and the ADB programme are verified by either the development banks or by the recipient.

Two main trade-offs can be identified as a result of the analysis. While a high level of recipient discretion and outcome orientation is generally desirable, this might increase the risk of gaming and distortions. Another trade-off exists, if direct indicators are unreliable or inconsistent over time. If this is the case, it has to be carefully weighted between using an unreliable direct indicator to avoid measure fixation and using a proxy indicator which is more reliable but may create adverse incentives.

5 Conclusions and recommendations

Based on the comparison of the five pilot programmes, five tentative recommendations can be formulated on how to design results-based programmes and select indicators. The recommendations are based on the cursory analysis presented in this paper and help to design 'ideal-type' results-based programmes. In reality, however, there will be additional considerations beyond the criteria and trade-offs described in this analysis. Most significantly, the practice of implementing results-based approaches will have to confront the tension between theoretical quality concerns and practical concerns relating to the specific country context, including costs and feasibility for instance. Practitioners need to carefully consider the theoretical soundness and practical considerations, when selecting indicators. Ultimately, any choice of indicators should be based as much as possible on the country context of the specific development intervention. In addition, the following recommendations could provide a useful supplement.

41 The length of the contract period is less an issue in the Twaweza intervention because here teachers are directly targeted. It can be assumed that there are several measures that teachers can directly take to improve student learning outcomes. This is especially the case if weak student performance may be correlated with high rates of teacher absenteeism and low teacher motivation, as in Tanzania.

42 Although there is to date no evidence of adverse effects, there is evidence that the contract period may be too short. The evaluations of the DFID Rwanda and Ethiopia pilot programmes both show that measured increases in results to date were not due to the results-based approach, and that both governments did not undertake substantial, additional activities to achieve the result.

1) Results focus: results-based approaches should as much as possible rely on outcome-indicators for making disbursements

Donors and partner countries should try to be as innovative as possible and rely strongly on outcome indicators. In cases where pre-financing is needed or where little certainty on the results-chain exists, a gradual approach with stronger reliance on process indicators in the beginning could be applied. The objective, however, should be to shift towards outcome-indicators as soon as possible.

2) Control: results-based approaches should emphasise results orientation over attribution concerns

The incentivised actor should have plausible control over achieving the activities and results that are measured by disbursement-linked indicators. It is not necessary to solely rely on process and output indicators in results-based approaches just to establish attribution. It should, however, be possible to demonstrate that the results-based approach is one of the main causes of observed changes in indicators. Outcome indicators should primarily be used to guarantee strong results orientation, while making sure that they can be influenced by the incentivised actor within the programme time frame.

3a) Financial incentives: results-based approaches should put more emphasis on leverage effects than on value for money

When deciding how to price results, funders should put more emphasis on potential leverage effects and risks for partners than on value-for-money. If value-for-money is the main criterion, a significantly reduced incentive effect might be the consequence. As a results, a programme may be efficient (low input for outcome) but not very effective in terms of achieving desired objectives.

3b) Financial incentives: results-based approaches should not set targets but rather reward incremental improvements

Linking disbursements to the achievement of certain thresholds is not advisable because it is extremely difficult to set targets that are both ambitious and realistically achievable. Targets that are set too low or too high do not create strong performance incentives. Therefore, results-based approaches should disburse for incremental improvements, in order to create stable incentives. When disbursing for incremental improvements, it is important that funders refrain from setting annual ceilings for indicators because this creates implicit targets and harmful expectations.

4) Measurability and verification: results-based approaches should rely on national systems and administrative data for measuring results, complemented by independent verification

Donors and partner countries should refrain from setting up parallel monitoring and verification systems that duplicate a country's own systems. This increases costs and undermines ownership. Instead, donors should rely on administrative data and commission independent verification to guarantee impartiality in determining disbursements.

5) Unintended consequences: the best mitigation strategy for avoiding unintended consequences is to use direct indicators and to ensure a robust verification process

Results-based approaches, just as any other aid intervention, can lead to unintended consequences that must already be taken into account at the design stage. To minimise unintended effects, it is important that results-based approaches are only implemented if good measures exist, i.e. if indicators are direct measures of the indented result. Proxy indicators should only be used if it is clear that that there is an overlap between actions needed to achieve the indicator result and actions needed to achieve the underlying objective. Moreover, to avoid distortions it has to be ensured that the partner is able to significantly influence results with genuine efforts only and that sufficient financing is available. Independent verification of results is also crucial to avoid data manipulation. Transparency and non-financial incentives, such as public peer pressure, can be important complements in this regard.

Bibliography

Adam, C. S., & Gunning, J. W. (2002). Redesigning the aid contract: Donors' use of performance indicators in Uganda. *World Development, 30*(12), 2045-2056.

Aman, K. D., & Balozi, M. A. (2014). Constructing theory with Big Result Now (BRN): Contribution of Tanzania to the theory of knowledge sharing behaviour. *International Journal for Innovation Education and Research, 2*(08), 7-21.

Asian Development Bank. (2013a). *Education sector development program (2013-2017): Disbursement linked results and indicators*. Manila: Author.

Asian Development Bank. (2013b). *Piloting results-based lending for programs* (Policy Paper). Manila: Author.

Asian Development Bank. (2013c). *Proposed results-based loans Democratic Socialist Republic of Sri Lanka: Education sector development program*. Manila: Author.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration, 84*(3), 517-538. doi: 10.1111/j.1467-9299.2006.00600.x

Binnendijk, A. (2000). *Results based management in the development co-operation agencies: A review of experiences*. OECD/DAC Working Party on Aid Evaluation.

Birdsall, N. (2013). *A critical moment for COD Aid or "the trouble with targets".* Retrieved from http://www.cgdev.org/blog/critical-moment-cod-aid-or-%E2%80%9C-trouble-targets%E2%80%9D

Birdsall, N., & Perakis, R. (2012). *Cash on delivery aid: Implementation of a pilot in Ethiopia.* Washington, DC: Center for Global Development.

Birdsall, N., & Savedoff, W. (2010). *Cash on delivery: A new approach to foreign aid*. Washington, DC: Center for Global Development.

Birdsall, N., Savedoff, W. D., Mahgoub, A., & Vyborny, K. (2011). *Cash on delivery: A new approach to foreign aid*. Washington, DC: Center for Global Development.

Busan Partnership. (2011). *Busan partnership for effective development co-operation: Fourth high level forum on aid effectiveness*, Busan. Retrieved from http://www.oecd.org/development/ effectiveness/ 49650173.pdf

Chee, G., Hsi, N., Carlson, K., Chankova, S., & Taylor, P. (2007). *Evaluation of the first five years of GAVI immunization services support funding*. Bethesda, MD: Abt Associates Inc.

Clist, P., & Verschoor, A. (2014). *The conceptual basis of payment by results*. Norwich, UK: University of East Anglia.

COSTECH, EDI, JPAL-MIT, & Twaweza. (2012). *Improving literacy and numeracy in Tanzania: Tanzania randomized evaluation in education 2013/4.* Dar es Salaam, Tanzania: Twaweza.

Danida. (2006). *Monitoring at programme and project level: General issues*. Copenhagen: Ministry of Foreign Affairs of Denmark.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1).

Delorme, P., & Chatelain, O. (2011). *Policy steering: The role and use of performance measurement indicators: Aid Delivery Methods Programme.* Retrieved from http://www.dochas.ie/Shared/Files/4/ Guide_on_Performance_Measurement.pdf

Department for International Development. (2011a). *Independent verification and evaluation of a pilot project of results based aid in the education sector in Ethiopia: Business case*. London: Author.

Department for International Development. (2011b). *Pilot project of results based aid in the education sector in Ethiopia: Business case*. London: Author.

Department for International Development. (2011c). *Rwanda education sector programme 2011/12 to 2014/15 business case*. London: Author.

Department for International Development. (2011d). *UK aid: Changing lives, delivering results*. London: Author.

Department for International Development. (2013a). *Big results now – business case and intervention summary*. London: Author.

Department for International Development. (2013b). *First annual review: Pilot project of results-based aid in Ethiopia*. London: Author.

Department for International Development. (2013c). *Second annual review: Pilot project of results-based aid in Ethiopia*. London: Author.

Department for International Development. (2014a). *Intervention summary – big results now in Education Results Based Programme, Tanzania (2014-2018)*. London: Author.

Department for International Development. (2014b). *Operational Plan 2011-2016*. DFID Ethiopia. London: Author.

Doran, G., Miller, A., & Cunningham, J. (1981). *There's a S.M.A.R.T. way to write management's goals and objectives*. *Management Review, 70*(11), 35-36.

Energy Sector Management Assistance Program. (2013). *Results-based financing in the energy sector: An analytical guide*. Washington, DC: ESMAP – World Bank.

Federal Democratic Republic of Ethiopia. (2010). *Education Sector Development Program IV (ESDP IV)*. Addis Ababa: Ministry of Education.

Gelb, A., & Hashmi, N. (2014). *Anatomy of PforR* (Working Paper 374). Washington, DC: Center for Global Development.

Health & Education Advice & Resource Team. (2013). *Independent verification of educational data for a pilot of results-based aid (RBA) in Rwanda – baseline report*. Oxford, United Kingdom: Author.

Holzapfel, S. (2014). *The role of indicators in development cooperation: An overview study with a special focus on the use of key and standard indicators* (Studies 81). Bonn: German Development Institute/ Deutsches Institut für Entwicklungspolitik (DIE).

Janus, H. (2014). *Real innovation or second-best solution? First experiences from results-based aid for fiscal decentralisation in Ghana and Tanzania* (Discussion Paper 3/2014). Bonn: German Development Institute/Deutsches Institut für Entwicklungspolitik (DIE).

Jonesa, S., Schipper, Y., Rutod, S., & Rajanib, R. (2014). Can your child read and count? Measuring learning outcomes in East Africa. *Journal of African Economies, 0*(0), 1-30. doi:10.1093/jae/eju009

Kapur, D., & Whittle, D. (2010). Can the privatization of foreign aid enhance accountability? *International Law and Politics, 42*, 1143-1180.

Keijzer, N., & Janus, H. (2014). *Linking results-based aid and capacity development support: Conceptual and practical challenges* (Discussion Paper 25/2014). Bonn: German Development Institute/Deutsches Institut für Entwicklungspolitik (DIE).

Klingebiel, S. (2012). *Results-based aid (RBA): New aid approaches, limitations and the application to promote good governance* (Discussion Paper 14/2012). Bonn: German Development Institute/ Deutsches Institut für Entwicklungspolitik (DIE).

Klingebiel, S., & Janus, H. (2014). Results-based aid: Potential and limits of an innovative modality in development cooperation. *International Development Policy, 6*.

Koeberle, S., Stavresk, Z., & Walliser, J. (Eds.). (2006). *Budget support as more effective aid? Recent experiences and emerging lessons*. Washington, DC: The World Bank.

Lockheed, M. E. (2008). *Measuring progress with tests of learning: Pros and cons for "cash on delivery aid" in education* (Working Paper 147). Washington, DC: Center for Global Development.

Manda, C., & Schipper, Y. (2014). *Kiufunza: Improving early grade learning outcomes.* Presentation held at the technical workshop on results-based funding, 19-20 May, OECD Conference Centre, Paris. Retrieved from http://www.oecd.org/dac/peer-reviews/Improving-Early-Grade-Learning-Outcomes.pdf

MDF Training & Consultancy. (2005). *MDF tool: indicators.* Retrieved from http://www.toolkitsportdevelopment.org/html/resources/40/408CC56F-509A-40D8-BE46-D7EEB4261F97/10%20Indicators.pdf

Ministry of Education Sri Lanka. (2013). *Education Sector Development Framework and Programme (ESDFP) (2013-2017).* 'Isurupaya', Battaramulla, Sri Lanka: Ministry of Education.

Musker, P., Clist, P., Abbott, P., Boyd, C., & Latimer, K. (2014). *Evaluation of Results Based Aid in Rwandan Education: 2013 evaluation report*. London: Upper Quartile and Institute of Policy Analysis and Research Rwanda.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257.

O'Brien, T., & Kanbur, R. (2013). *The operational dimension of results based financing*. Retrieved from http://www.kanbur.aem.cornell.edu/papers/OperationalDimensionsOfRBF.pdf

O'Brien, T., Fiszbein, A., Gelb, A., Kanbur, R., & Newman, J. (2012). *Results based development assistance: Perspectives from the South Asia region of the World Bank* (Working Paper). Ithaca, New York: Charles H. Dyson School of Applied Economics and Management, Cornell University.

O'Donnell, O., van Doorslaer, E., Wagstaff, A., & Lindelow, M. (2008). *Analyzing health equity using household survey data: A guide to techniques and their implementation* (World Bank Institute Learning Resources Series). Washington, DC: The World Bank.

Organisation for Economic Co-operation and Development. (1993). *OECD core set of indicators for environmental performance reviews*. Paris: Author.

Organisation for Economic Co-operation and Development. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: Author.

Organisation for Economic Co-operation and Development/Development Assistance Committee. (2009). *Glossary of key terms in evaluation and results based management: Glossar entwicklungspolitischer Schlüsselbegriffe aus den Bereichen Evaluierung und ergebnisorientiertes Management*. Paris: Author.

Pearson, M., Johnson, M., & Ellison, R. (2010). *Review of major results based aid (RBA) and results based financing (RBF) schemes.* Retrieved from http://www.oecd.org/dac/peer-reviews/Review-of-Major-RBA-and-RBF-Schemes.pdf

Perakis, R., & Savedoff, W. (2015). *Does results-based aid change anything? The roles of pecuniary interests, visibility, accountability and discretion in four case studies* (CGD Policy Paper 052). Washington, DC: Center for Global Development.

Perrin, B. (2013). *Evaluation of payment by results (PBR): Current approaches, future needs* (Working Paper 39). London: Department for International Development.

Prennushi, G., Rubio, G., & Subbarao, K. (2002). Monitoring and evaluation. In J. Klugman (Ed.), *A sourcebook for poverty reduction strategies: Volume 1: Core techniques and cross-cutting issues* (pp. 105-130). Washington, DC: The World Bank.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*(2), 247-252.

Sandefur, J., & Glassman, A. (2014). *The political economy of bad data: Evidence from African survey & administrative statistics* (Working Paper 373). Washington, DC: Center for Global Development.

Savedoff, W. D., & Martel, K. D. (2011). *Cash on delivery aid for health: What indicators would work best?* (Working Paper 275). Washington, DC: Center for Global Development.

Shepard, L. A. (1991). *Will national tests improve student learning?* Paper presented at the American Educational Research Association Public Interest Invitational Conference, Accountability as a state reform instrument: Impact on teaching, learning, and minority issues and incentives for improvement. Washington, DC.

Smith, P. (1995). Of the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration, 18*(2-3), 277-310.

Twaweza. (2014). *2014 Annual Plan*. Dar es Salaam, Tanzania: Author.

University of California, Los Angeles. (2013). *What is the difference between categorical, ordinal and interval variables?* Retrieved from http://www.ats.ucla.edu/stat/mult_pkg/whatstat/nominal_ordinal_interval.htm

United Nations Development Group. (2011). *Results-based management handbook: Harmonizing RBM concepts and approaches for improved development results at country level.* Retrieved from http://www.ilo.org/public/english/bureau/program/dwcp/download/undg_rbm1011.pdf

United Nations Development Programme. (2002). *RBM in UNDP: Selecting indicators.* Retrieved from http://www.undp.org/eo/documents/methodology/rbm/indicators-Paperl.doc

United Nations Development Programme. (2009). *Handbook on planning, monitoring and evaluating for development results*. New York, NY: Author.

United States Agency for International Development. (2010). *Performance monitoring & evaluation TIPS: Selecting performance indicators* (2nd ed.). Retrieved from http://pdf.usaid.gov/pdf_docs/pnadw106.pdf

Uwezo. (2014). Uwezo website. Retrieved from http://www.uwezo.net/

Vogel, I. (2012). *Review of the use of 'Theory of Change' in international development.* Report commissioned by the Department for International Development. Draft – review report and practical resources. London: Department for International Development (DFID).

World Bank. (1996). *Performance monitoring indicators handbook* (World Bank Technical Paper No. 334). Washington, DC: Author.

World Bank. (2012). *Program appraisal document for urban local government strengthening program (Program-for-Results Operation).* Washington, DC: Author.

World Bank. (2014). *Tanzania – Big Results Now in Education Program Project, Program Appraisal Document*. Washington, DC: World Bank Group.

World Bank and African Economic Research Consortium. (2012). *Service delivery indicators education and health: Tanzania.* Washington, DC: The World Bank.

Annex

Annex Table 1: World Bank "Big Results Now in Education Program" Tanzania				
Objective (results chain level)	**Indicator**	**Typology (I-Input, A-Activity, O-Output, OO-Outcome, D-Direct, I-Indirect, QT-Quantitative, QL-Qualitative)**	**Scale or threshold**	**Value (million USD)**
DLI#1: Recipient has completed all the Foundational Activities (Activity)	DLI#1.1: The Recipient's Ministry of Education and Vocational Training (MoEVT), Prime Minister's Office, Regional Administration and Local Government (PMO-RALG) and Ministry of Finance agreed with the Association and Co-financiers on a 4-year Big Results Now Education (BRNEd) Budget Framework, coded against specific budget lines in the Recipient's Budget	A, D, QL	Threshold	3.1
	DLI#1.2: The Recipient's MoEVT has reviewed and revised (i) Education Management Information System (EMIS) staff job descriptions; aligned with appropriate civil service grading system; and (ii) finalised and approved school census format and guidelines and disseminated to all district education officers in FY2013/14	A, D, QL	Threshold	3.1
	DLI#1.3: The Recipient's MoEVT and PMO-RALG have agreed on a format for measuring and collecting verifiable data related to the transfer of Capitation Grants to schools in FY2013/14	A, D, QL	Threshold	3.1
	DLI#1.4: The Recipient's MoEVT has prepared a list of all primary and secondary schools on the Recipient's territory with information on location, number of students and teachers in FY2013/14	A, D, QL	Threshold	3.1
	DLI#1.5: The Recipient's MoEVT has prepared a format for (i) measuring and collecting verifiable data related to primary pupil-teacher ratio (PTR) at school and LGA level and disseminated a summary primary PTR report format, with guidance notes to all Local Government Authority (LGAs) (including acceptable ranges between 35- 50); and (ii) established baseline values for 2013-14 and, agreed upon with PMO-RALG, 4 year targets for primary school PTRs in FY2013/14	A, D, QL	Threshold	3.1

Annex Table 1 (cont.): World Bank "Big Results Now in Education Program" Tanzania				
Objective (results chain level)	**Indicator**	**Typology (I-Input, A-Activity, O-Output, OO-Outcome, D-Direct, I-Indirect, QT-Quantitative, QL-Qualitative)**	**Scale or threshold**	**Value (million USD)**
DLI#2: Recipient has evidenced timely and adequate resource flows for the Program (Input)	DLI#2.1: The Recipient has released, quarterly, total levels of funds as per BRNEd Budget Framework	I, D, QT	Scale (amount allocated is pro-rated in terms of percentage of funding released)	31
	DLI#2.2: The Recipient has quarterly released full amount of Capitation Grants (CG) to schools within each LGA	I, D, QT	Scale (amount allocated is pro-rated in terms of percentage of funding released)	6
DLI#3: The Recipient has produced an Annual Summary Education Performance Report (ASEPR) and an annual school-level EMIS data by each LGA (Activity)	DLI#3.1: The Recipient has released an Annual Summary Education Performance Report in acceptable format;	A, D, QL	Scale (number of LGAs for which data exists)	6
	DLI#3.2: The Recipient has made available online an annual school-level EMIS data	A, D, QL	Scale (percentage of schools for which data is available online)	6

Annex Table 1 (cont.): World Bank "Big Results Now in Education Program" Tanzania				
Objective (results chain level)	**Indicator**	**Typology (I-Input, A-Activity, O-Output, OO-Outcome, D-Direct, I-Indirect, QT-Quantitative, QL-Qualitative)**	**Scale or threshold**	**Value (million USD)**
DLI#.4: Teachers have been deployed efficiently across and within districts (Output)	DLI#4.1: The Recipient has met the annual target for number of LGAs achieving the acceptable range for primary PTRs	O, D, QT	Threshold and scale (number of LGAs achieving the acceptable range for primary PTRs as compared to target)	7.7
	DLI#4.2: The Recipient has met the annual target for number of primary schools achieving the acceptable range of primary PTRs	O, D, QT	Scale (number of schools achieving the acceptable range for primary PTRs as compared to target)	12.3
DLI#5: Number of schools receiving school incentive grant (SIG) as indicated in the Program design (Activity/Input)	DLI#5.1: The SIG and student teacher enrichment plan (STEP) Guidelines have been prepared	A, P, QL	Threshold	6.2
	DLI#5.2: The Recipient has met the annual target for number of schools that have received SIG grant	I, D, QT	Scale (number of schools receiving SIG per annum as compared to target)	6
	DLI#5.3: The Recipient has met the annual target for number of schools that have conducted STEP activities	A, P, QT	Scale (number of schools conducting STEP activities per annum as compared to target)	6

Annex Table 1 (cont.): World Bank "Big Results Now in Education Program" Tanzania				
Objective (results chain level)	**Indicator**	**Typology (I-Input, A-Activity, O-Output, OO-Outcome, D-Direct, I-Indirect, QT-Quantitative, QL-Qualitative)**	**Scale or threshold**	**Value (million USD)**
DLI#6: Recipient has demonstrated an increase in student learning outcomes (Outcome)	DLI#6.1: The Recipient's MoEVT has developed and agreed upon the Reading, Writing, and Arithmetic (3R) assessment tools with the PMO-RALG in FY2013/14	A, P, QL	Threshold	3.3
	DLI#6.2: The Recipient has met the annual target of improvement in words per minute (wpm) in national 3R average FY 2014/2015	OO, D, QT	Threshold and scale (3.9 million are pro-rated in terms of improve-ment in wpm)	16
Source: World Bank (2014)				

Annex Table 2: ADB "Education Sector Development Programme" Sri Lanka				
Objective (results chain level)	**Indicator**	**Typology**	**Scale or threshold**	**Value (million USD)**
DLR 1: Pass rates for General Certificate of Education (GCE) 'O' level examinations increased (Outcome)	June 2016: The pass rate for GCE 'O' Levels increased to at least 63%.	OO, D, QT	Threshold	10
	June 2017: The pass rate for GCE 'O' Levels increased to at least 65%.			
DLR 2: Pass rates for GCE 'A' level examinations increased (Outcome)	June 2016: The pass rate for GCE 'A' Levels increased to at least 64%.	OO, D, QT	Threshold	10
	June 2017: The pass rate for GCE 'A' Levels increased to at least 64%.			
DLR 3: Pathways from school to Technical and Vocational Education and Training (TVET) developed – Technology Stream commenced and implemented at GCE 'A' levels (Activity and Output)	June 2013: MOE notifies policy decision to introduce a Technology Stream at GCE 'A' levels in accordance with the ESDFP (2013-2017)	A, D, QL	Threshold	8
	June 2014: Detailed time-bound plans for implementation of the Technology Stream finalized as per approved standards, i.e., addressing curricula for the Technology Stream, gender-sensitive teacher deployment and training, and teaching resources and facilities required clearly defined and approved by Ministry of Education (MOE)	A, D, QL	Threshold	8
	June 2015: Public awareness campaigns completed by MOE to promote enrolment in the Technology Stream, including promoting the enrolment of girls in the Technology Stream	O, D, QL	Threshold	8
	June 2016: Technology Stream piloted at GCE 'A' levels in at least 100 schools, with trained teachers, at least 50% of whom are women, and adequate teaching resources and facilities according to the standards set for the 1000 Schools Development Program	O, D, QT	Threshold	4
	June 2017: Assessment of initial 100-school pilot and a time-bound plan for expansion and gender-inclusiveness of the Technology Stream to other schools completed	A, D, QL	Threshold	4
DLR 4: Secondary schools upgraded to offer all subject streams (Output)	June 2013: At least 250 Type 1C schools (offering Arts and Commerce GCE 'A' levels streams only) identified, including at least 10 provincial schools in each province and 5 national schools in each district by MOE for upgrading, including infrastructure upgrading, to Type 1AB status following the approved standards of the 1000 Schools Development Program	A, P, QT	Threshold	8
	June 2014: a) Time-bound development plans, including provision for facilities for girls (toilets) and computer-assisted learning, for the 250 selected schools completed and submitted to MOE. b) Upgrading initiated in at least 50 % of selected schools as per standards of the 1000 Schools Development Program	a) A, P, QL b) A, P, QT	a) Threshold b) Threshold	8
	June 2015: a) Upgrading initiated in at least 90% and b) completed in at least 50% of the 250 selected schools as per standards of the 1000 Schools Development Program	a) A, P, QT b) O, D, QT	a) Threshold b) Threshold	8

Annex Table 2 (cont.): ADB "Education Sector Development Programme" Sri Lanka				
Objective (results chain level)	**Indicator**	**Typology**	**Scale or threshold**	**Value (million USD)**
DLR 4: Secondary schools upgraded to offer all subject streams (Output) (cont.)	June 2016: At least 50% of the 250 selected schools upgraded to Type 1AB status (offering all streams), with at least 25% female enrolment in the Science Stream and Technology Stream as per approved standards.	O, D, QT	Threshold	8
	June 2017: At least 85% of the 250 selected schools upgraded to Type 1AB status (offering all streams), with at least 25% female enrolment in the Science Stream and Technology Stream as per approved standards.			
DLR 5: Enrolment in GCE 'A' levels Science Stream increased (Outcome)	June 2016: Enrolment in GCE 'A' levels Science Stream increases to 26% (at least 23% for girls)	OO, D, QT	Threshold	10
	June 2017: Enrolment in GCE 'A' levels Science Stream increases to 28% (at least 25% for girls)			
DLR 6: Enrolment in GCE 'A' levels Commerce stream increased (Outcome)	June 2016: Enrolment in GCE 'A' levels Commerce Stream increases to 30% (at least 26% for girls)	OO, D, QT	Threshold	10
	June 2017: Enrolment in GCE 'A' levels Commerce Stream increases to 33% (at least 29% for girls)			
DLR 7: Principals and deputy principals trained (Output)	June 2013: Training modules for school principal and deputy principal training are developed by Center for Education Leadership Development (CELD) and approved by MOE, and their implementation approved.	A, P, QL	Threshold	8
	June 2014: At least (cumulative) 200 principals (25% women) and 300 deputy principals (30% women) complete training as per approved standards.	O, D, QT	Threshold	24
	June 2015: At least (cumulative) 300 principals (at least 25% women) and 500 deputy principals (at least 30% women) complete training program at CELD as per approved standards.			
	June 2016: At least (cumulative) 500 principals (at least 25% women) and at least 700 deputy principals (at least 30% women) complete training as per approved standards.			
	June 2017: At least (cumulative) 600 principals (at least 25% women) and (cumulative) 900 deputy principals (at least 30% women) complete training program at CELD as per approved standards.			
DLR8: Institutional capacity at Ministry of Education and provincial levels strengthened (Output)	June 2013: Sector Monitoring and Technical Support Unit (SMTSU) established with Program Head and key staff assigned.	O, P, QL	Threshold	8
	June 2014: a) Performance-based partnership agreements agreed signed among MOE and at least 7 provinces, reflecting alignment between central and provincial annual work and budget plan (AWBPs) for Education Sector Development Framework and Program (ESDFP) and b) indicating that at least 80% of the initial capital budget for FY2013 for school education has been spent.	a) A, P, QL b) I, P, QT	a) Threshold b) Threshold	8

Annex Table 2 (cont.):	**ADB "Education Sector Development Programme" Sri Lanka**			
Objective (results chain level)	**Indicator**	**Typology**	**Scale or threshold**	**Value (million USD)**
DLR8: Institutional capacity at Ministry of Education and provincial levels strengthened (Output) (cont.)	June 2015: a) Performance-based partnership agreements reflecting alignment between central and provincial AWBPs for ESDFP updated based on review of previous year progress of AWBP and b) indicating that at least 80% of the initial capital budget for school education has been spent.	a) A, P, QL b) I, P, QT	a) Threshold b) Threshold	16
	June 2016: a) Performance-based partnership agreements reflecting alignment between central and provincial AWBPs for ESDFP updated based on review of previous year progress of AWBP and b) indicating that at least 85% of the initial capital budget for FY2015 for school education has been spent.	a) A, P, QL b) I, P, QT	a) Threshold b) Threshold	16
	June 2017: a) Performance-based partnership agreements reflecting alignment between central and provincial AWBPs for ESDFP updated based on review of previous year progress of AWBP and b) indicating that at least 85% of the initial capital budget for FY2016 for school education has been spent.			
DLR 9: Improved transparent and efficient procurement (Output)	June 2013: a) Online system for posting of procurement information including bid notices, contract awards, complaints information etc. is developed and b) all information is easily accessible to public free of cost.	a) O, D, QL b) O, D, QL	a) Threshold b) Threshold	8
	June 2014: Sample bidding documents are developed for ESDFP and procurement procedures are included in the program operational manual for ESDFP by February 2014.	A, P, QL	Threshold	8
	June 2015: The findings of the procurement audit conducted as part of the ADB"s Annual Fiduciary Review (AFR) show no more than 50% of audited contracts have substantial negative findings by third year of operation.			
	June 2016: The findings of the procurement audit conducted as part of the ADB"s AFR show no more than 40% of audited contracts have substantial negative findings by fourth year of operation.			
	June 2017: The findings of the procurement audit conducted as part of the ADB"s AFR show no more than 30% of audited contracts have substantial negative findings by fifth year of operation.			
Source: ADB (2013a)				

Annex Table 3: DFID "Results-Based Aid Pilot - Education Sector Programme" Rwanda				
Objective (results chain level)	**Indicator**	**Typology**	**Scale or threshold**	**Value (million USD)**
Equitable access to quality education and improved learning outcomes (Outcome)	Additional children sitting the P6 exam	OO, D, QT	Scale	14
	Additional sitting the S3 exam	OO, D, QT	Scale	
	Additional sitting the S6 exam	OO, D, QT	Scale	
Qualified, suitably-skilled and motivated teachers and trainers to meet demands of expanding education access (Output)	Additional teachers reaching an agreed level of English language competency by 2014 in comparison to a baseline in 2012	O, D, QT	Scale	
Source: DFID (2011); HEART (2013)				

Annex Table 4: DFID "Pilot Project of Results Based Aid in the Education Sector" Ethiopia				
Objective (results chain level)	**Indicator**	**Typology**	**Scale or threshold**	**Value (million USD)**
Improved access to, and quality of, lower secondary schooling (Outcome)	Additional students sitting the national grade 10 examination	OO, D, QT	Scale	47
	Additional students passing the national grade 10 examination	OO, D, QT	Scale	
Source: DFID (2011b)				

Annex Table 5: Twaweza – "KiuFunza Local Cash on Delivery" Tanzania				
Objective (results chain level)	**Indicator**	**Typology**	**Scale or threshold**	**Value (million USD)**
Improve learning outcomes in primary schools (Outcome)	No. of children who pass the proven Uwezo literacy (English) assessment at the end of the school year	OO, D, QT	Scale	0.59
	No. of children who pass the proven Uwezo literacy (Kiswahili) assessment at the end of the school year	OO, D, QT	Scale	
	No. of children who pass the proven Uwezo numeracy (Mathematics) assessment at the end of the school year	OO, D, QT	Scale	
Source: COSTECH et al. (2012)				

Publications of the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE)

Studies

87 Ashoff, Guido. (2015). *Die Global Governance-Qualität der internationalen Aid Effectiveness Agenda: Eine theoretische Analyse und Bewertung der Systemreform der internationalen Entwicklungszusammenarbeit* (220 pp.). ISBN 978-3-88985-663-0.

86 Hampel-Milagrosa, Aimée. (2014). *Micro and small enterprise upgrading in the Philippines: The role of the entrepreneur, enterprise, networks and business environment* (169 pp.). ISBN 978-3-88985-640-1.

85 Zelli, Fariborz, Daniela Erler, Sina Frank, Jonas Hein, Hannes Hotz, & Anna-Maria Santa Cruz Melgarejo. (2014). *Reducing emissions from deforestation and forest degradation (REDD) in Peru: A challenge to social inclusion and multi-level governance* (178 pp.). ISBN 978-3-88985-651-7.

84 Fues, Thomas, & Jiang Ye (Eds.). (2014). *The United Nations Post-2015 Agenda for Global Development: Perspectives from China and Europe* (373 pp.). ISBN 978-3-88985-649-4.

83 Horstmann, Britta, & Günther Schulz-Heiss. (2014). *Providing international adaptation finance for vulnerable communities: A study on potentials and limits of social investment funds* (106 pp.). ISBN 978-3-88985-647-0.

[Price: EUR 10.00; publications may be ordered from the DIE or through bookshops.]

Discussion Papers

1/2015 Gerardo Bracho. (2015). *In search of a narrative for Southern providers: The challenge of the emerging economies to the development cooperation agenda* (52 pp.) ISBN 978-3-88985-664-7.

37/2014 Büge, Max. (2014). *Do preferential trade agreements increase their members' foreign direct investment?* (54 pp.). ISBN 978-3-88985-655-5.

36/2014 Meijer, Karen. (2014). *Can supply chain initiatives reduce deforestation? A comparative analysis of cases from Brazil and Indonesia* (38 pp.). ISBN 978-3-88985-662-3.

35/2014 Brandi, Clara, & Max Büge. (2014). *A cartography of the new middle classes in developing and emerging countries* (33 pp.). ISBN 978-3-88985-661-6.

34/2014 Chan, Sander, & Pieter Pauw. (2014). *A global framework for climate action: Orchestrating non-state and subnational initiatives for more effective global climate governance.* (41 pp.). ISBN 978-3-88985-660-9.

[Price: EUR 6.00; publications may be ordered from the DIE or through bookshops.]

For a complete list of DIE publications:
www.die-gdi.de