EQUIP-Tanzania Impact Evaluation

Midline Technical Report, Volume II

Methods and Supplementary Evidence

FINAL REPORT

Authors: Georgina Rawle, Nicola Ruddle, Gunilla Pettersson Gelander, Johanna Wallin, Michele Binci, Paul Jasper, Jana Harb, Madhumitha Hebbar, Jean Davis and Alice Aldinucci.

March 2017

Acknowledgements

Please consult Volume I of this report to see a full list of acknowledgments.

Oxford Policy Management Limited	Level 3, Clarendon House	Tel	+44 (0) 1865 207 300
	52 Cornmarket St	Fax	+44 (0) 1865 207 301
	Oxford OX1 3HJ	Email	admin@opml.co.uk
Registered in England: 3122495	United Kingdom	Website	www.opml.co.uk

Table of contents

List of figures and tables

List of abbreviations

ADEM	Agency for the Development of Education Management
ATT	Average treatment effect on the treated
BL	Baseline
BRN-Ed	Big Results Now in Education
CAPI	Computer-assisted personal interviewing
CF	Community facilitator
COSTECH	Tanzania Commission for Science and Technology
CSO	Civil society organisation
CTAs	Community Teaching Assistants
DEO	District Education Officer
DFID	Department for International Development (UK)
EGMA	Early Grade Maths Assessment
EGRA	Early Grade Reading Assessment
EMIS	Education management information system
EQUIP-T	Education Quality Improvement Programme in Tanzania
ESDP	Education Sector Development Programme
ETP	Education and Training Policy
FGD	Focus group discussion
GoT	Government of Tanzania
IE	Impact evaluation
IGA	Income-generating activity
INSET	In-service training
KII	Key informant interview
LANES	Literacy and Numeracy Education Support Programme
LGA	Local Government Authority
MA	Managing Agent
MDAs	Ministries, departments and agencies
ML	Midline

MoEST	Ministry of Education, Science and Technology
MoEVT	Ministry of Education and Vocational Training
NECTA	National Examinations Council of Tanzania
NSLCF	National School Leadership Competencies Framework
OPM	Oxford Policy Management
PFM	Public financial management
PO-RALG	President's Office Regional Administration and Local Government
PMO-RALG	Prime Minister's Office Regional Administration and Local Government
PEDP	Primary Education Development Programme
PSA	Programme support activity
PSLE	Primary School Leaving Examination
PSM	Propensity score matching
PSM–DID	Propensity score matching–differences in differences
PTP	Parent–Teacher Partnership
REO	Regional Education Officer
RG	Reference Group
RTI	Research Triangle International
RTL	Regional team leader
SC	School committee
SEDP	Secondary Education Development Programme
SEQAS	Specialist Evaluation and Quality Assurance Services
SIS	School Information System
SLM	School leadership and management
SRP	School Readiness Programme
TA	Technical assistance
TCF	Teacher Competency Framework
TDNA	Teacher Development Needs Assessment
TIE	Tanzania Institute of Education
TOC	Theory of change

TOR	Terms of reference
TTC	Teacher training college
TZS	Tanzanian shilling
UCL	University College London
UNESCO	UN Educational, Scientific and Cultural Organization
UNICEF	UN Children's Fund
USDM	University of Dar es Salaam
VTF	Village task force
WEC	Ward Education Coordinator
WSDP	Whole-School Development Plan

Part D Introduction

1 Introduction

1.1 Overview

This is the second Volume of the midline (ML) impact evaluation (IE) report of the Education Quality Improvement Programme in Tanzania (EQUIP-T). EQUIP-T is a four-year, Government of Tanzania (GoT) programme funded by the UK Department for International Development (DFID). It targets seven of the most educationally disadvantaged regions in Tanzania in order to increase the quality of primary education and improve pupil learning outcomes, in particular for girls. It is a large programme – costing approximately £50 million and expected to reach about 2.3 million pupils.

The IE ML research aims to assess the impact, effectiveness and cost of EQUIP-T so far, approximately 20 months into implementation, and to draw out implications for programme adjustment and consolidation.

The ML IE report is organised into two volumes. Volume I (Results and Discussion), presents the main findings and draws recommendations and lessons for the programme's managing agent (MA) and the GoT to consider in seeking to strengthen the programme. Volume II (Methods and Supplementary Evidence) contains technical methods sections, as well as more detailed qualitative findings and supplementary quantitative analysis to support the conclusions reached in Volume I. Readers interested in the more in-depth evidence base for the ML findings should read both Volumes.

1.2 Structure of this Volume

Volume I contains three parts: Part A: objectives, background and methods; Part B: findings; Part: C implications and conclusion. Volume II is divided into three further parts, as follows:

- **Part D: introduction**: this short overview (Chapter 1).
- **Part E: methods**: this comprises four chapters each covering a different element of the mixed methodology. Full details of the overall IE design and methods are given in Volume II of the baseline (BL) IE report (OPM 2015b). This detail is not repeated here, but instead the key features are summarised and, where relevant, updated to include changes made at ML as well as the specific risks and limitations to the ML analysis. The four chapters cover: mixed methods (Chapter 2); quantitative design (Chapter 3); qualitative design (Chapter 4); and costing methods (Chapter 5).
- **Part F: evidence:** these six chapters presents the more detailed quantitative and qualitative evidence which supports the main findings in Volume I. The chapters cover: impact estimation (Chapter 6)–this gives a technical overview of the techniques applied to the ML data and discusses the results in greater depth than was possible in Volume I; supplementary descriptive quantitative analysis (Chapter 7)–this describes additional trends in key indicators in programme treatment schools; and qualitative thematic research on teacher capacity and performance (Chapter 8), school leadership and management (SLM) (Chapter 9), district planning and management (Chapter 10) and community engagement (Chapter 11).

These three parts are supplemented by 11 annexes, which include: processes for IE stakeholder engagement and governance (Annex B); ML fieldwork details (Annexes C, D and I), statistical tables of results from the programme treatment districts (Annex F), a technical annex on the measurement of pupil learning (Annex G), and a table mapping parts of the ML IE report to the Specialist Evaluation and Quality Assurance Services (SEQAS) evaluation criteria (Annex K).

Part E Methods

2 Mixed methods approach

2.1 Overview

The key purposes of the IE study are to: i) generate evidence on the impact of EQUIP-T on primary pupil learning outcomes, including any differential effects for boys and girls; ii) examine perceptions of effectiveness of different EQUIP-T components; iii) provide evidence on the fiscal affordability of scaling up EQUIP-T post 2018; and iv) communicate evidence generated by the IE to policy-makers and key education stakeholders.

The IE uses a mixed methods approach where quantitative and qualitative methods are integrated to provide robustness, depth and improved validity of the research findings. This rests on integrating methodologies for better measurement, sequencing information for better data collection, and (iteratively) merging findings for better action (Carvalho and White 1997; Gabarino and Holland 2009). The IE BL report notes the 14 practical steps taken in the IE which frame the mixed methods approach, including, for example 'using quantitative data to highlight priority issues and generate new hypotheses to test through qualitative research' and 'using qualitative data in a subsample of quantitative areas to elicit rich contextual data that might affect programme outcomes' (See OPM 2015b, p2 for more details). The specific actions taken during the ML research to mix methods are outlined below.

Figure 1 illustrates the IE data collection design, and shows the relationship between quantitative and qualitative components. The quantitative survey collects data at school level covering head teachers (HTs), teachers and pupils in a representative sample of programme treatment districts, while the qualitative research takes place in small purposive sample of schools (and their communities), wards, districts, regions and also at national level. The rationale for the different focus of the research components rests on the relative strength of each method in capturing different types of changes–this is explained further in relation to the ML research design below. The methods are complementary to allow for both robust measurement of programme impact, as well as deeper insights into the potential mechanisms through which the programme may have influenced the outputs and outcomes of interest.

Figure 1: **Relationship between quantitative and qualitative data collection**

QUANTITATIVE	LEVEL	QUALITATIVE
	Sample of regions	KIIs with EQUIP-T staff in regional offices & regional education officials
	Sample of districts	KIIs with district education officials
	Samples of wards	Key informant interviews with ward education officers
Head teacher interview	Sample of schools	KIIs with head teacher & community leader
Teacher interview, classroom observations, teacher assessments	Sample of teachers	FDG with teachers
Pupil test and interview with parents	Sample of pupils	FGDs with pupils
	Sample of community members	FGDs with school committees , fathers and mothers (separately)

Source: OPM impact evaluation team. Note: KII means key informant interview, and FDG means focus group discussion

The IE design has three main components. The first is the quantitative estimation of impact, which uses quantitative data from the BL and ML school surveys (this is explained in greater detail in Chapters 3 and 6). The second is the 'rigorous factual analysis' to explain programme impact.[1] This mixes evidence from the quantitative survey and the qualitative research in programme treatment areas, in order to understand key channels of programme influence, or reasons for ineffectiveness, using research questions structured around the TOC (these are set out in the ML evaluation matrix–the development of this framework is explained below). The third is the costing study, which is intended to better understand the costs of the programme and what this would mean for the affordability of continuing or scaling up the programme (or parts of the programme) after EQUIP-T comes to an end.

2.2   Mixed methods approach at ML

Members of the quantitative and qualitative research teams worked together at the ML design, data collection and analytical stages of this ML IE study to ensure that the study is able to take full advantage of its mixed methods design. The main processes at each stage are described below.

Mixed methods at ML design stage

The research priorities for the ML IE are captured in a comprehensive ML evaluation matrix (see Volume I Annex B). This sets out evaluation questions linked to the programme TOC, and identifies sources of evidence to answer each question—either the quantitative survey or qualitative research or both. The ML evaluation matrix was jointly developed by the lead quantitative and qualitative researchers in January/February 2016. To establish ML research

---

[1] The term 'rigorous factual analysis' comes from White (2009).

priorities, the integrated team reviewed BL findings and programme implementation documents, and then visited Tanzania to interview a range of key education sector stakeholders, including EQUIP-T programme staff (see OPM, 2016a for more details of this process). Following this, the ML tools for both methods were developed concurrently with the quantitative and qualitative researchers working closely together to ensure complementarity.

As at the BL, the research areas emphasised in the qualitative and quantitative research differ because of their different methodological strengths. The ML qualitative research focuses particularly on district planning and management, community participation and demand for accountability, teacher motivation and the relationship between teachers, HTs and WECs, because qualitative methods are well-suited to capture perceptions of changing roles, responsibilities, and relationships. By contrast, the quantitative research is suitable for robustly documenting and quantifying changes in pupil learning, teacher capacity and performance, and school leadership and management, while still covering some quantifiable aspects of community engagement with schools and district planning and management, and to estimate programme impact on pupil learning and selected other outcomes.

Mixed methods at ML data collection stage

The concurrent timing of the ML quantitative survey and qualitative research (April/May 2016), whereby qualitative research teams visited schools shortly after the quantitative survey team had a number of advantages for strengthening the research process. Observations on research priorities from the early stages of the quantitative fieldwork were shared with the lead qualitative researchers which enabled them to modify parts of the planned research. For example, a decision was taken to include non-participatory lesson observation in each of the qualitative research sites to enrich the quantitative evidence. It was also clear from the early quantitative evidence that the ML research would benefit from further qualitative focus on understanding reasons for changes in classroom absenteeism. The qualitative team were also able to contact the school in advance of visits using updated information from the quantitative team, in order to arrange for the participants to be available at the school.

Mixed methods at ML analysis stage

An early sharing of draft quantitative and qualitative findings during a team workshop held in early August 2016 permitted a rich discussion and pointed to areas for further investigation in both data sets. For example, in the case study schools, seven out of nine HTs were new since the BL and many of the changes in SLM observed in the case study schools were attributed by respondents to this change. This prompted further exploration of the HT turnover estimates from the quantitative survey data to try to establish more about the recent career and training history of HTs, so as to be able to gauge more clearly the programme's potential influence. The district- and ward-level qualitative data was also re-examined to explore themes related to head teacher and teacher transfers.

This type of information sharing and enrichment continued into the report writing phase. Each of the thematic findings chapters in Volume I was co-authored by a member from each of the quantitative and qualitative teams. This 'buddy' system works by sharing and commenting on iterative drafts, and though this process continuing to explore both datasets, for example to validate or confute links in the programme TOC supported by one set of data. Draft recommendations and lessons drawn from the research were strengthened through iterative discussions and reflection from both the quantitative and qualitative teams.

3 Quantitative design and ML adjustments

The core IE quantitative methodology involves measuring a consistent set of indicators in panel of 200 schools (100 treatment and 100 control schools) over three rounds of research (2014, 2016 and 2018), conducted at the same time of year (April/May). The core quasi-experimental methods used at BL were replicated at ML, but some adjustments were necessary to the instruments, school-level sampling and to the fieldwork model, and these are explained below. The BL IE report volume II (OPM 2015b, Chapter 3, pp2-28) explains the full details of the design including the rationale, sampling strategy, instruments, impact estimation methods, as well as key methodological risks and limitations. This detail is not repeated here, but the core design features are summarised briefly in the sections below, together with an explanation of the adjustments made for the ML research, a table showing ML sample outturn, and consideration of specific risks, and limits to the quantitative component at ML.

3.1 Rationale for quasi-experimental design

The main purpose of the IE is to be able to robustly attribute changes in key impact-level and outcome-level indicators to EQUIP-T as a whole. The EQUIP-T Managing Agent (MA) purposively selected by the regions and districts into the programme on the basis of these being disadvantaged in terms of education and other social and economic indicators. In the absence of random assignment, a pure randomised control trial (RCT) was not possible, and the IE employed the best possible approach to simulate the RCT approach. In this case, this was to mimic randomisation using propensity score matching (PSM), and then to employ a PSM and difference-in-difference (DID) approach to estimating programme impact (see Chapter 6 for details of how the impact estimation was carried out in practice using BL and ML data).

3.2 Sampling strategy and sample size

3.2.1 Sampling strategy

Prior to sampling, a list of eligible treatment and control districts was established by excluding districts that are: (i) in Lindi and Mara as these are part of the EQUIP-T programme but not covered by the IE; (ii) receiving other education programmes that aim to influence the same outcomes as EQUIP-T including Big Results Now in Education (BRN-Ed), Kiufunza, UNICEF's school-based INSET programme and USAID's TZ21 programme; (iii) part of OPM's BL pre-tests.

The sampling was carried out in four stages:

1. **Selection of control districts**: PSM was used to match eligible control districts to the 17 pre-selected eligible treatment districts.
2. **Selection of treatment schools**: schools in the treatment districts were selected using stratified random sampling.
3. **Selection of control schools**: PSM was used to match eligible control schools to the sample of treatment schools.
4. **Selection of pupils and teachers within schools**: pupils and teachers were sampled within schools using systematic random sampling. The within-school sampling was assisted by selection tables automatically generated within the computer assisted survey instruments. Information on the detailed procedure followed by enumerators to select the 15 pupils, is in Box 1 . Table 1 shows the ML survey respondents, sampling and instruments. One change was made to the BL sampling approach—instead of sampling teachers of Standards 1 to 3 for interview, all of them were interviewed at ML (see table footnote for details). The aim was to

boost the sample size, given BL information that most schools had three Standard 1-3 teachers or fewer.

Box 1 Detailed procedure for within-school pupil sampling

Enumerators were trained to use the following procedure for sampling Standard 3 pupils:

Collect the Standard 3 attendance registers for all streams, and check that they are filled in for 'today', then follow this sampling procedure:

1. Use a pencil. Have a rubber available.
2. Starting at 1, write a sequential series of numbers beside the names of all pupils who are present today.
3. If there is more than one stream in Standard 3, continue the number series on to the next registers.
4. The final number in your pencil number series is the number of pupils present today in Standard 3. Enter this number into the cell in the computer assisted personal interviewing (CAPI) instrument.
5. The CAPI instrument will automatically produce 15 pupil selection numbers in red font.[1]
6. Look again at the pencil number series you marked on the register/s. Find the pupil name which corresponds to the first selection number. Write the pupil's name into the sampled pupils table.
7. Repeat the step above for the other 14 selection numbers. You will need to scroll in the table to see the spaces to enter all of the sampled pupils.[2]

Source: OPM (Midline Fieldwork Manual, April 2016). Notes: 1) The CAPI instrument automatically generates a random set of 15 different numbers of maximum value equal to the number in Step 4. 2) There is also a procedure for replacement in the event that any of the 15 pupils cannot take the test for a valid reason, for example being ill.

Table 1: ML IE quantitative survey respondents, sampling and instruments

Respondent	School-level Sample	Instrument
Standard 3 pupils	Sample (15 pupils present on the day)	Adapted Early Grade Reading Assessment (EGRA) Adapted Early Grade Maths Assessment (EGMA) Pupil background
Parents of tested Standard 3 pupils	Sample (15 parents)	Poverty score card
Standards 1 to 3 Kiswahili and maths teachers[1]	No sampling	Interview
Standards 1 to 3 Kiswahili teachers	Sample (up to 3) Treatment schools only	Teacher Development Needs Assessment Kiswahili (TDNA Kiswahili)
Standards 1 to 3 mathematics teachers Standards 4 to 7 mathematics teachers	Sample (up to 3 from each group) Treatment schools only	Teacher Development Needs Assessment mathematics (TDNA mathematics)
Head teacher	No sample	Interview School Records
Enumerator observation	No sample Treatment schools only	Lesson observation
Enumerator observation	No sample	Head count (of teacher and pupil attendance)
Source: OPM ML survey. Note: (1) At BL, a sample of Standards 1-3 teachers were interviewed (all Standards 1-3 teachers were sampled to take a TDNA were also interviewed).		

The sampling strategy yields a panel of schools (same schools visited during each round), and a repeated cross-section of Standard 3 pupils and Standards 1-3 teachers.

At BL five out of 200 schools had to be replaced during fieldwork using a carefully controlled reserve list. Replacement was not necessary at ML, and all schools that were interviewed at BL were revisited and interviewed at ML (see Annex D for ML fieldwork details).

3.2.2   Sample size

The theoretical justification for the choice of target sample sizes for each unit is explained in the BL report volume II (p8).

Actual response rates are high in the ML survey, as they were for the BL survey. Table 2 shows that actual sample sizes at ML are close to target sample sizes. For tested Standard 3 pupils and their parents, response rates are almost 100%. The response rate drops to 93% for Standards 1 to 3 teacher interviews, including 8% of teachers who were absent or unavailable on the day and later interviewed by phone. Response rates for teachers taking teacher development needs assessments (TDNAs) are slightly lower at around 85%; one of the reasons being that it was sometimes difficult for teachers who teach both maths and Kiswahili to spare time to take both TDNAs.[2] The target for lesson observations was 200, but under the new Standards 1 and 2 curriculum, maths (arithmetic) and Kiswahili (either reading or writing) lessons often run sequentially without a break and this enabled 94 maths lessons to be observed and 137 Kiswahili lessons, more than the target.

Table 2:   ML quantitative survey actual and intended sample sizes

Sampling unit	Treatment sample			Control sample		
	Target sample	Actual sample	Actual/Target (%)	Target sample	Actual sample	Actual/Target (%)
Regions	5	5	100	7	7	100
Districts	17	17	100	8	8	100
Schools	100	100	100	100	100	100
Std. 3 pupils (tested both in Kiswahili and maths)[1]	1484	1483	99.9	1488	1488	100
Scorecards[1]	1484	1477	99.5	1488	1486	99.9
Std. 1-3 Kiswahili/maths teacher interviews[2]	434	405	93.3	422	412	97.6
Std. 1-3 Kiswahili TDNAs	283	243	85.9	n.a.	n.a.	n.a.
Std. 1-3 Maths TDNAs	285	239	83.9	n.a.	n.a.	n.a.
Std. 4-7 Maths TDNAs	270	231	85.6	n.a.	n.a.	n.a.
Std. 2 lesson observation maths[3]	100	94	94.0	n.a.	n.a.	n.a.
Std. 2 lesson observation Kiswahili[3]	100	137	137	n.a.	n.a.	n.a.

Source: IE ML survey. Notes: (1) In 4 treatment schools and 4 control schools, there were fewer than 15 eligible pupils, so the targets are fewer than 1500. (2) The samples includes 21 HTs/acting HTs (treatment) and 14 (control) who teacher Stds 1-3. Out of the 39 teachers in treatment and control schools who did not sit for the interview, one refused while 38 were unavailable (absent on the day and couldn't be reached over the phone later). Some 11% of teachers in treatment schools and 6% in control schools were interviewed over the phone. (3) There were 94 maths (arithmetic) lessons and 137 Kiswahili lessons (either reading or writing)

---

[2] Each TDNA takes up to one hour to complete.

observed. Some of these subjects were taught consecutively (without a break) in one class period. There were 172 separate class periods observed.

3.2.3   Survey weights

In order to obtain indicator estimates that are representative of the EQUIP-T programme areas (more specifically, the 17 districts that comprise the IE sample), to feed into the descriptive trends analysis of the TOC, estimates were weighted using survey weights. The survey weights are computed as the normalised values of the inverse probabilities of selection into the sample for each unit of observation. The formulae for computing the weights for different units (schools, pupils and teachers) are in the BL IE report volume II (p11). At BL there were two sets of teacher weights, one for sampled teachers (those who were interviewed or took TDNAs) and one for roster teachers (for indicators which use data on all teachers in a school). At ML this was extended to three sets of teachers weights because all Standard 1-3 teachers were interviewed rather than a sample.

The survey weights were applied within a survey set up in Stata (the statistical programme used to analyse the data) that takes account clustered sampling, stratification and finite population corrections.

3.3   Survey instruments

The ML survey round uses a set of survey instruments that retain most of the BL questions, but with some additions to take into account changes in programme context and design, and focus of programme implementation. Table 3 summarises the instruments that were used at BL. The enumerators administered all of the instruments using Computer Aided Personal Interviewing (CAPI), except for the TDNA which were administered on paper, as these take the form of mock pupil tests which teachers mark.

Table 3:   Quantitative survey instruments from BL

Description of content	
1. Standard 3 pupil Kiswahili test (same pupils tested in both Kiswahili and mathematics)	
• **Kiswahili literacy pupil test based on standard 1 and 2 curriculum requirements** • **Pupil background information**	Early Grade Reading Assessment (EGRA) Pupil background
2. Standard 3 pupil mathematics test (same pupils tested in both Kiswahili and mathematics)	
• **Mathematics pupil test based on standard 1 and 2 curriculum requirements**	Early Grade Mathematics Assessment (EGMA)
3. Parents of Standard 3 tested pupil interview	
• **Set of household characteristics (that can be used to convert scores into poverty likelihoods based on a pre-existing instrument)**	Poverty score card
4. Standards 1, 2 and 3 teacher interview	
• **Background information: gender, age, years of teaching, qualifications** • **Frequency/type of in-service training received** • **Frequency/nature of lesson observation and nature of feedback** • **Frequency/nature of performance appraisal**	Teacher interview
5. Standards 1, 2 and 3 teacher development needs assessment Kiswahili	

• **Teacher Kiswahili subject knowledge assessment based on the primary school curriculum (standards 1-7 but only limited materials from standards 1 and 2)**	Teacher Development Needs Assessment Kiswahili (TDNA)
6. Standards 1, 2 and 3 teacher development needs assessment mathematics	
• **Teacher mathematics subject knowledge assessment based on the primary school curriculum (standards 1-7 but only limited materials from standards 1 and 2)**	Teacher Development Needs Assessment mathematics (TDNA)
7. Standards 4-7 teacher development needs assessment mathematics	
• **Teacher mathematics subject knowledge assessment based on the primary school curriculum (standards 1-7 but only limited materials from standards 1 and 2)**	Teacher Development Needs Assessment mathematics (TDNA)
8. Head teacher interview and data collection from school records	
• **Background information on head teacher: gender, age, years of experience, qualifications** • **School background information: teachers, physical facilities, school timetable, number of days school open** • **Frequency/type of school planning/management in-service training received** • **Teacher attendance (by records)** • **Pupil attendance (by records)**	Head teacher interview School records checks
9. Standard 2 Kiswahili and mathematics lesson observations	
• **Inclusive behaviour of teachers with respect to gender and seating position** • **Key teacher behaviours in the classroom** • **Availability of lesson plan** • **Availability of seating, textbooks, exercise books, pens/pencils etc. during the lesson**	Lesson observation
10. Headcount observation	
• **Pupil and teacher attendance observation** • **Physical facilities observation**	Head count
Source: Adapted from OPM, 2015b, pp12-13.	

It is critical that the core information collected over multiple rounds remains the same, so that key indicators can be reliably tracked over time. This means that changes to the BL instruments are not substantial. The main changes to instruments are:

- **Pupil background**: addition of questions on pre-school attendance; languages spoken at home and use of languages by teachers at school.

- **Poverty score card**: addition of questions on languages spoken at home; support at home for homework and learning to read; child work.

- **Teacher interview**: addition of questions on languages spoken at home and at school; more details on INSET (including capturing courses provided by LANES and other programmes) and views on EQUIP-T INSET if applicable; teaching the new curriculum; teaching practices related to inclusion; views on HTs and parent-teacher-partnership (PTP) actions on school improvement; journey time to school; timeliness and accuracy of salary payments; reasons for school and classroom absenteeism.

- **Head teacher interview**: addition of questions on languages spoken at home; more details on inservice training (INSET) and views on EQUIP-T INSET if applicable; monthly reporting to the WEC/district; views on the PTP; own actions to improve the school; actions resulting from a community needs assessment if applicable; reasons for teacher school and classroom

absence; actions taken by WECs during visits; timeliness and accuracy of salary payments; reasons for own absenteeism.

- **School records collection**: addition of questions on pre-school classes and provision in community (including SRP); background data on all teachers in school (age, experience, qualifications) not just the sampled teachers; more details on material and financial resources received by the school; attendance records for sampled pupils. Removal of detailed questions on the school timetable because this was time-consuming at BL and the data was not comprehensive enough to be useful, and there were concerns about its reliability.

- **Lesson observation**: adaption to account for 3Rs lessons where reading, writing and maths are taught sequentially as one 'lesson'; addition of questions on teacher's approach to teaching reading and whether materials such as those supplied by EQUIP-T (supplementary readers, Big Books, Teacher Read-Alouds) are in the classrooms and being used by teachers or pupils.

Overall the addition of questions is greater than removal and so this has increased the time required in each school to collect the data. To accommodate this, there have been some changes to the fieldwork model (see next section).

The revisions to the BL instruments, and the fieldwork model, were trialled during two ML pre-tests which took place in November 2015 and February 2016 (see Annex D).

The original development of the instruments, and their contents, is described in detail in the IE BL report volume II (pp13-18). Given the importance of the measurement of pupil learning to the IE (improving learning achievement is the main goal of the programme), it is worth briefly summarising the test design process. The OPM design team worked with a national team of specialists comprising Kiswahili and maths specialists from the University of Dar es Salaam, primary school teachers and a Tanzanian test design specialist, to develop the two pupil tests. The team developed new items adapted from an existing Early Grade Reading Assessment (EGRA) and an Early Grade Maths Assessment (EGMA) that was being used to monitor the Government's BRN-Ed programme. Three pre-tests with purposive sampling were carried out to check item difficulty and discrimination, clarity of wording, protocols for accurate measurement, and child-friendliness. The test items are kept secure so that they can be re-administered each round.

3.4 Fieldwork timing and model

From a measurement perspective, it is important that the data collection for each round takes place at the same time of year, so that seasonal influences on measurements are kept to a minimum. This has largely been achieved for the BL and ML rounds which both took place during April and May, although the ML fieldwork was a few weeks later than BL because the school holidays and examination timetable fell on different dates each year.

The ML fieldwork uses a slightly different model to BL, partly because the fieldwork window was shorter. Instead of a BL model where teams of three visited schools over two days, the ML model uses teams of five or six enumerators to visit schools in one day. This had implications for the approach to interviewing parents. At BL parents were requested to visit the school to be interviewed on the second day of fieldwork. Under the ML model, enumerators tracked pupils' households to conduct interviews with parents or guardians there. For the small minority of cases where this was impossible, interviews were conducted over the phone. The overall response rate was close to 100%. The one-day model, increased the risk of Standard 1-3 teachers being absent for interview but this was mitigated using phone interviews for absent teachers (see notes under Table 2), and response rates were well over 90%.

More details on fieldwork are summarised in Annex D (and full details are available in a separate fieldwork report, OPM 2016d). This covers fieldwork personnel and selection criteria; preparation (including details of the two pre-tests); training and piloting; fieldwork organisation and execution, quality control and data checking protocols, and fieldwork challenges and lessons for the endline.

3.5 Quantitative analysis

3.5.1 Impact estimation

The quasi-experimental design relies on PSM and DID techniques to estimate programme impact on a small set of key impact indicators. As noted in the BL IE volume II (OPM 2015b, p24) the first stage of this approach requires the pupil and teacher samples used to generate the impact indicators to be balanced across treatment and control groups. Following the BL, further technical work was undertaken with the aim of achieving sample balance which partly depends on sample sizes achieved in the field (see OPM 2015c for more details). The results of this analysis were reported in the Midline Planning Report (OPM 2016a) as follows:

- given the balance achieved across covariates of matched treatments and controls for the pupil samples, a quantitative estimation of impact can be confidently undertaken;
- some of the BL information on pupils' background characteristics was missing (age) or was not deemed adequate (language), with a negative effect on the estimation model;
- the samples of teachers who took the TDNA are too small to be able to achieve a satisfactory balance across treatment and controls; and
- the samples of teacher interviewed are probably large enough to be able to achieve a satisfactory balance across treatment and controls, but there is insufficient background information for model construction.

These results were used to adjust the ML data collection: i) to collect additional data to obtain: missing age data from pupils tested at BL; age data for ML pupils from multiple sources; richer information on pupils' language background; basic background information on *all* teachers in the school (age, experience, qualifications) as well as for teachers who have left the school since BL. ii) to conduct TDNAs and lesson observations in programme schools only, because in light of the results above, the information from control schools could not be usefully used in quantitative impact estimation. The small sample size risk for teachers was noted in the baseline report volume II (OPM 2015b p19).

Chapter 6 contains full technical details on the estimation of programme impact on the following indicators (including further details on methodology and robustness checks, and the full results):

EQUIP-T impact: Better learning outcomes, especially for girls

- Proportion of Standard 3 pupils in the bottom Kiswahili performance band (%)
- Proportion of Standard 3 pupils in the top Kiswahili performance band (%)
- Proportion of Standard 3 pupils in the bottom mathematics performance band (%)
- Proportion of Standard 3 pupils in the top mathematics performance band (%)

This allows assessment of (i) changes in the proportion of pupils moving from the lowest performance band to any of the higher bands and (ii) changes in the proportion of pupils moving into the top performance band from any of the lower bands. Annex G on the measurement of pupil learning outcomes explains the how the performance bands are linked to curriculum-based

competencies. Chapter 2 in Volume I disaggregates these indicators by gender and assesses how learning gaps have changed between BL and ML.

EQUIP-T outputs 1 and 2: Enhanced professional capacity and performance of teachers and enhanced school leadership and management skills

- Teacher school absenteeism (%)
- Teacher classroom absenteeism (%)
- Proportion of Standard 1-3 teachers who report participation in performance appraisal (%)

Results from the estimation of programme impact on three additional indicators of school leadership and management, were not reliable enough to report with confidence, given concerns about potential measurement errors in the indicators (see Section 3.6 below on limitations). The indicators affected are: proportion of Standards 1-3 teachers that: (i) could not show any evidence of assessing pupil academic progress; (ii) who report lesson observations by head teacher in last 30 days: (iii) report head teacher feedback on lesson plans in the last 30 days. These indicators have been reported for programme schools with caveats in the descriptive trend analysis.

3.5.2   Descriptive trend analysis in programme schools

Descriptive trend analysis of key indicators related to the programme TOC feeds into the 'rigorous factual analysis' (see Chapter 2) which seeks to understand why changes have happened or not as anticipated in the TOC. As explained in Chapter 2 Section 0 above, the quantitative and qualitative descriptive findings are integrated in an iterative analysis process.

3.5.3   Quality assurance

The impact estimation analysis was reviewed internally by OPM's statistical methods team, and then by a UK-based academic researcher, familiar with these methods and their application in education. Two other reviewers (a senior Tanzanian academic and an ex-World Bank Senior Education Specialist for Tanzania), provided comments and feedback on the descriptive analysis and interpretation. Annex B provides further details on the overall quality assurance processes applied in this study.

3.6   Risks and limitations

3.6.1   Contamination risk from other programmes

As the BL IE report volume II (p26) highlights, the most common risk in longitudinal surveys is potential contamination of the selected impact study areas by third party interventions that may affect the outcomes of interest to the evaluation.

In this case, the main risk of contamination, also highlighted at BL, comes from the government's Literacy and Numeracy Support Programme (LANES) because it planned to operate in areas which included the IE control districts. The IE team highlighted this risk prominently in the BL IE report, discussed this with senior government officials, DFID and representatives of the LANES programme (including providing written notes explaining this risk) during November 2014 meetings, and followed up with subsequent correspondence. The firm view of government stakeholders was that implementation of LANES must be national from an equity perspective, and that staggered roll-out was not practical given the design. Stakeholders also jointly held the view that the initial

LANES interventions were unlikely to pose a serious contamination risk, as the likely impact on pupil learning would be minimal (too dilute) prior to the ML round of the IE, without further inputs.

The ML Planning Report (OPM, 2016a) followed up on LANES activities since BL and confirms that LANES does include the IE control districts. There are two main LANES activities which affect the control schools: (i) two Standard 1 and 2 teachers were invited to a ten day centralised INSET orientation on the new 3Rs curriculum; and (ii) HTs were invited to a three day regional INSET on school leadership and management. The ML IE survey collected data on LANES implementation in the control districts so that the IE could take account of the contamination risk. Box 2 explains how the impact identification strategy is still able to robustly estimate the treatment effect of EQUIP-T.

<div style="border:1px solid #000;">

Box 2 Dealing with the contamination risk from the LANES programme in impact estimation

The rigorous identification of programme impact relies on a comparison between a set of treatment and control schools (a credible counterfactual). One of the conditions for a credible counterfactual is that there is no contamination of outcome measures, i.e. that no other interventions interfere with the outcomes of interest in control areas. Otherwise, the true treatment effect is confounded and cannot be identified.

As explained above, since the EQUIP-T BL research, the LANES programme has implemented INSET for early grade teachers and HTs in areas which include the IE control districts, with the aim of improving early grade learning outcomes.

The impact identification strategy used in this ML IE study can still robustly estimate the treatment effect of EQUIP-T for three reasons. First, given the centralised modality of one-off INSET delivery and the limited number of participants from each school, it is reasonable to assume that the extent of contamination of outcome measures is likely to be fairly low (OPM, 2016a). Second, the IE ML has dealt with the remaining risk by collecting survey data on LANES INSET in the control schools, which confirms that coverage of the LANES INSET across the control schools is very high and uniform across schools. Finally, this means that the impact estimates presented in this report can be interpreted as the impact of EQUIP-T compared to a counterfactual situation without EQUIP-T but with LANES INSET equivalent training. The main implication for the EQUIP-T impact estimation is therefore the need for careful interpretation of the impact estimates (see Chapter 6 for further details on interpretation of impact estimates).

</div>

Another programme, Big Results Now in Education (BRN-Ed) started before EQUIP-T and the IE BL sampling excluded the 60 districts affected by school-level BRN-Ed activities. As noted in the ML IE planning report (OPM 2016a, p10), recent BRN-Ed activities directed at early-grade pupils, their teachers and HTs, have been implemented via LANES in a partnership so pose no additional contamination risk to the EQUIP-T IE. A review of other large education programmes operating in Tanzania (OPM 2016a, p11) did not reveal any additional contamination risks, and the ML survey data did not pick up any substantial programme activity from third parties relevant to the evaluation.

The IE team will continue to monitor the implementation of LANES, and other programmes over the next year before endline, and discuss with DFID and the IE Reference Group if there is a more serious contamination risk that cannot be dealt with in the manner outlined in the box above. A higher risk of contamination may derive from a change in the nature and frequency of the INSET delivery. If the LANES programme replicates the INSET orientation for both teachers and head teachers multiple times and increases the length of the sessions and the number of participants, this would be likely to have a more substantial effect on outcome measures. In this case, the additional marginal impact of EQUIP-T over LANES INSET would be reduced and our ability to detect this impact with statistical confidence would be undermined. Similarly, if additional LANES INSET activities are not distributed across our control area but systematically target only some specific control schools, this would affect the uniformity of the control group. The overall comparison between treatment and control groups could still be performed as the outcome indicators are averaged at the group level. However, we would have to run a sensitivity analysis across our control schools to determine whether any systematic difference in LANES distribution is

driving and therefore skewing the indicator values in the control group. Careful and continuous monitoring of LANES operational plans and roll-out will provide us with an indication of whether the currently contained risk of contamination will have increased by EL.

If the contamination risk becomes untenable by EL, then it would be necessary to substantially alter the design of the IE in order to meet its objectives given the evolving context. The most obvious possibility would be to use a more extensive theory-based attribution approach to assessing programme impact. This would have re-design implications for both the quantitative and qualitative work, and possibly the sampling. For example, the list of outcome indicators measured by the quantitative survey in programme treatment districts would almost certainly need to be expanded and adjusted in line with more of the detailed causal pathways in the TOC. Obviously the nature of the information on impact provided by this type of methodology is different to the quantification and attribution of gains in learning outcomes to a programme based on a quasi-experimental approach which uses a counterfactual.[3]

3.6.2 Confounding TOC failure with implementation failure

This is the risk that the programme does not have the expected effect because it is not implemented as intended, rather than because its underlying design and TOC is flawed. For example, key activities don't happen or happen too slowly or that different activities happen. The design of the IE aims to address this by using the TOC to frame the analysis (see ML evaluation matrix in Volume I Annex B) in order to understand if changes have happened as anticipated or not, and to explain why. It is important to note, however, that the IE does not include a process evaluation so its focus is on how effective the programme has been in meeting its objectives, rather than on the details of receipt of inputs and timing of activities. The IE relies on information from the EQUIP-T Managing Agent (MA) in its annual reports on the status of implementation. This documentation is comprehensive and gives a good overview of the implementation of different components overall, and notes any major adjustments to programme design (which can in turn be used by the IE team in designing each round of research). However, school- ward- or district-level implementation data is not readily available on component activities, and so a detailed analysis of the status of implementation in the areas under evaluation is not possible.[4]

3.6.3 Limitations to the quantitative component

Table 4 describes general limitations of the quantitative component and mitigating factors at ML. Some of the specific limitations to the ML quantitative analysis include:

- High item non-response rates in a few areas: although generally item response was generally high (see Annex F which contains statistical tables with actual sample sizes for all indicators presented in Volume I. Target sample sizes are in Table 2 in Chapter 3), the following areas had higher rates of non-response and thus have a greater risk of bias in the estimates:

    Receipt of capitation grant: about 15% of schools did not have complete records on capitation grants received in the previous two school years. As at BL (and despite some simplification of the instrument), accurate data is difficult to collect. The financial records for capitation grants are held in different formats, making it difficult to establish the precise timing of grant receipt, and sometimes head teachers (particularly those new to the job) don't understand

---

[3] This type of substantial redesign of the final round of the IE would be likely to require additional resources.

[4] During interviews with EQUIP-T staff in January 2016, it was clear that the programme is making efforts to strengthen it management information system for capturing the implementation of activities. This means that for endline round of the IE, more detailed data may be available on activities which take place between ML and endline.

them well. The capitation grant results presented in Volume I are only approximate and are heavily caveated.

Head teacher interview questions that could not be asked over the phone: although phone interviews were conducted with most of the 15% of HTs that were absent on the day of the survey, some follow-up questions related to school records were not asked.

Teacher interview questions that could not be asked over the phone: although phone interviews were conducted with about 9% of teachers because of absence, it was not possible to ask them the questions which required them to show written records (examples of pupil assessment, feedback on lesson observation and lesson plans).

- TDNA sample lower than expected: the response rate for TDNAs at ML was about 85% for the three different samples (similar to BL rates, except for Standard 4-7 teachers where the response rate was 94%), partly because some sampled teachers were absent, but also because some teachers were selected for both TDNAs and an interview which was too time-consuming. A small percentage of teachers refused to do the TDNA at ML, possibly because knew about it from BL (see Annex D for more details). Thus there is some of risk of bias in the TDNA results.

- Problems with comparing BL and ML estimates of SLM indicators because of changes in administration: Teachers were asked to show written examples of their own pupil assessments, feedback on their lesson plans, and feedback on a lesson observed by the head teacher. At BL these examples were sought during the interview. During the ML pre-test this was observed to be very disruptive and so these questions were asked at the end of the interview. There is some suggestion from field feedback that some teachers were reluctant to look for evidence at the end of the interview because they wanted the interview to be finished. This may also have affected how they answered the previous questions on whether these actions had taken place (for example, answering 'Did you receive feedback from the head teacher on you lesson plans in the last 30 days?). It is difficult to unpick the possible effect of this change in administration, but it means that BL and ML results are not strictly comparable.

Table 4 Possible limitations to the quantitative component

Possible limitation	Why is this limiting and mitigating factors
EQUIP-T regions and districts were purposively selected to target those performing weakly on selected education indicators	An RCT design was not possible for the impact evaluation due to purposively selected treatment regions and districts. A quasi experimental PSM-DID approach was chosen instead to establish an appropriate counterfactual to assess EQUIP-T impact. This relies on the assumptions of PSM to mimic the experimental approach. A key assumption of PSM is that the information on observables is sufficient to match the control and treatment groups for the purposes of the evaluation. If the groups are matched on observables, but differ on unobservable, time-variant characteristics that are likely to affect the impact of the programme the estimate of impact would not be robust. *See Chapter 6 for further explanation of ways this risk has been minimised.*
Language spoken at home	Pupils that do not speak Kiswahili at home may be systematically disadvantaged by pupil tests conducted in Kiswahili. *Diagnostic tests on the ML pupil test data did not find any substantial differential item functioning related to home language.[1] The ML instruments also collected more information on language spoken at home and school, for use in the impact estimation.*
Not possible to substantially change survey instruments after the baseline	If there are substantial changes to the EQUIP-T programme design after the baseline the instruments may not be able to measure this. The indicators included have been carefully considered to ensure they capture key EQUIP-T outcome and outputs as per the original design. *The ML IE design was adjusted to accommodate some of the key changes in the EQUIP-T programme design such as providing supplementary reading materials to school, but there are limits to this. For example, the quantitative survey is not able to capture changes related to the new School Readiness Programme (although this was included to some extent in the qualitative research).*
The number of teachers per school is small in the control and treatment districts	This means that the total sample of teachers was smaller than originally anticipated with implications for the power of detection. A larger school sample size would have been required to address this issue but was not deemed possible by DFID for cost reasons. *See Section 3.5.1 for details on the effect of small teacher samples on the impact estimation.*
Relative short time period for assessing EQUIP-T impact on pupil learning	There has been less than two years between the school-level implementation of EQUIP-T and the ML IE survey. This is a relatively short time to expect any EQUIP-T impact on pupil learning, and impact may only be detectable by endline. *The results in Volume I show that it has been possible to detect programme impact between BL and ML.*
Inaccurate identification of EQUIP-T interventions by respondents	Respondents do not always know the official name of programme interventions or that they come from EQUIP-T. Multiple names for the same intervention may be in use, and there is the possibility of respondents mixing up EQUIP-T interventions with other similar development interventions. *The instruments were carefully pretested, and some questions were adjusted to deal with naming confusion which arose at this design stage. Similarly during enumerator training and piloting many school visits took place to practice the survey protocols and to review data collected. Daily debriefs were held, and changes to the instruments and training manual were made as appropriate. For example, in naming the type of INSET training teachers had received in 2015, some teachers talked about 3Rs training which our enumerators initially recorded as LANES 3Rs curriculum orientation training. On investigating this issue further, it emerged that the teachers were actually talking about EQUIP-T 3Rs curriculum related training. We trained enumerators to probe where the 3Rs training had taken place as we knew that the LANES 3Rs training all took place in Dodoma University.*

Notes: (1) The differential item functioning tests were carried out as part of a Rasch analysis of the pupil test data (see Annex G for information on measurement of pupil learning using the Rasch model).

4 Qualitative design and ML adjustments

This chapter describes the objectives of the qualitative research and how rigour is to be achieved through the qualitative research of the evaluation, including sampling, instruments, fieldwork and analysis.

4.1 Objectives of the qualitative research

While the quantitative component will allow rigorous attribution of changes in selected outcomes to the EQUIP-T programme as a whole, qualitative methods allow for the probing and exploration of the relationships between any such changes and the programme; exploration of changes in outcomes not amenable to quantification (for instance, understanding of responsibilities); and contextualisation of any such changes. The qualitative work principally focuses on the activity-output-outcome linkages, and also provides a basis to assess some of the assumptions in the programme's theory of change. It also allows an opportunity to identify any major changes not picked up in the quantitative component, as well as alternative reasons for change.

The qualitative methodology is informed by the objectives of the qualitative research: to assess hypotheses around the EQUIP-T programme's TOC, to elicit perceptions about the EQUIP-T programme, to explore impact hypotheses at levels not covered by the quantitative survey, and to investigate and further explore findings that arose from the baseline (BL). This range of objectives implies a qualitative methodology that is both confirmatory (assessing existing hypotheses) and exploratory (explaining impacts, developing new hypotheses and capturing unexpected impacts).

That said, the midline (ML) qualitative research differs slightly from BL, with less use of exploratory methods at ML. The focus groups and interviews at ML were more directed towards specific areas targeted through EQUIP-T, and how and why these have changed. For example, at BL head teachers were asked about the conditions of their school, whereas at ML they were asked what changes have taken place over the last two years, and their views on what led to these changes.

4.1.1 Rigour

A challenge in qualitative research is the definition and achievement of 'rigour', particularly, as in this case, when the research methodology should be open to the identification of new hypotheses, causes and unexpected impacts so should contain an emergent dimension not fully prescribed at the outset. Qualitative research is often accused of being 1) open to research bias or anecdotal impressions, 2) impossible to reproduce and 3) difficult to generalise (Mays and Pope 1995). OPM follows a protocol of ensuring rigour throughout the research by implementing specific strategies at each stage of the evaluation process – design, sampling, fieldwork, analysis and writing up. The main aim of these strategies is to minimise a single researcher bias and to be transparent in demonstrating the research process as well as data analysis. Throughout the following sections, the adopted strategies for ensuring rigour will be discussed as they relate to sampling and design, fieldwork and analysis.

4.1.2 Using theory to improve generalizability and inform structure

The TOC of the EQUIP-T programme – the sequence of events connected to EQUIP-T that is expected to lead to the EQUIP-T desired outcomes – reveals the expected causal pathways for programme impact, and is therefore extremely useful for generating hypotheses that structure the impact evaluation (Vogel 2012). It also helps address the larger policy question around whether the EQUIP-T model would work elsewhere in Tanzania (given that the EQUIP-T districts were not

randomly selected). If the impact evaluation finds that EQUIP-T worked in the programme districts, how would policymakers know whether it might work elsewhere? While there are at present no agreed best practice for how to answer this question (Cartwright and Hardie 2012), the impact evaluation uses the TOC to inform the structure of the analysis and to improve thematic generalisability.

The EQUIP-T TOC has been used to map out EQUIP-T's causal chain and contextual assumptions that must hold for EQUIP-T activities to lead to the desired impact (White 2009). The quantitative analysis will assess whether impact can be attributed to EQUIP-T, while (primarily) qualitative data will be used to conduct what White (2009) calls 'rigorous factual analysis' on whether the expected links in the causal chains hold (i.e. Why did this impact occur? Did this activity lead to this output, to this outcome, to this impact?), and whether the assumptions are valid, over time (i.e. is the context as we assumed it was?). This has been done by turning the existing TOC into an evaluation matrix (see Volume I Annex B) and then setting specific questions for (more) structured instruments and observations. As agreed in the study's TOR (see Volume I Annex A1.1.4), although the study is framed by the TOC, there are limits on the extent to which results can be generalised because it does not constitute a full theory-based evaluation. The 'rigorous factual analysis' focuses on selected causal pathways and assumptions that are considered by key stakeholders to be the most important, but does not cover the TOC comprehensively.

4.2   Design

The qualitative component at midline builds on the research model used in the baseline. For full details of the qualitative baseline design, see the baseline report volume II, (OPM 2015b, part F, Section 4 and Annex J).

4.2.1   Sampling of qualitative study sites

The sampling approach was theoretically informed and designed to generate responses from a selected number of individuals and groups that are broadly representative (though not statistically) of groups relevant to EQUIP-T, and which allow some identification of heterogeneous impact.[5]

During the baseline design process, qualitative methods were chosen to ensure depth of understanding of impact and to ask the 'why' questions. As such, in contrast to quantitative sampling, the focus was on gaining a deeper understanding of a smaller set of case studies rather than a wider breadth of data. Though qualitative findings are never representative in a statistical sense, the use of typical and extreme case sampling at baseline both on a district level and then within districts (see details below) allowed for findings to be thematically generalised and situated across a variety of contexts. This ensured that the heterogeneous contexts in which the programme operates were included as part of the qualitative sampling, whilst the fieldwork time and analytical focus given to each case study ensured that responses from a variety of individuals and groups were included for each case school. The choice of three districts and nine case study schools was judged to yield the best balance of depth of understanding of impact while still allowing for sufficient variation in context, given resources available for the qualitative research.

The same nine programme treatment schools and communities across three districts/regions that were sampled as sites for the qualitative research for the baseline were visited again at midline. The same sites were visited in order to retain the variation in context, to have greater information

---

[5] Note that only schools/districts/regions benefiting from the EQUIP-T programme were included in the qualitative study.

and understanding of the changes in outputs and outcomes in these sites, and to be able to follow up on the issues which arose in the baseline.

The baseline report gives a full description of how the nine sites were selected through purposive 'typical case sampling' and 'extreme case sampling'. Put simply this means sampling average, high and low performing districts and schools in terms of selected education inputs, outputs and outcomes. There were several stages to the sampling procedure. First, all districts in treatment areas were categorised as 'performing well, given school and pupil characteristics', 'performing typically, given school and pupil characteristics' and 'performing worse than expected, given school and pupil characteristics'. At the next stage, a district was chosen from each of the three categories based on the following criteria: (i) average, high and low absolute pupil performance had to be represented; (ii) a range of pupil and school social and economic resourcing had to be represented; and (iii) each district must be taken from a different region.

Following district selection, three schools within each of the three sampled districts were selected using purposive 'typical case sampling' and 'extreme case sampling' based on average, high and low academic performance relative to other schools in the selected district. In addition, the group of nine selected schools were deliberately chosen to ensure variation in the recent rate of academic improvement at baseline and in pupil to teacher ratios.

In order to preserve the confidentiality of respondents, the three districts/regions and nine schools are referred to by a label/number throughout both volume I and volume II of this report.

4.2.2   Instruments and methods

As with the BL, the qualitative part of the IE makes use of two research instruments – key informant interviews (KIIs) and focus group discussions (FGDs). All of the KIIs and FGDs utilised structured and unstructured methodologies. Structured methods allow for the efficient assessment of pre-specified hypotheses and unstructured methodologies allow for unanticipated or context-specifics to be captured and for new hypotheses to be developed.

The KII tools are thus semi-structured by design: particular themes already identified as of interest to the evaluation are assessed and open-ended questions are also posed. This flexibility allows the team to probe further and develop inquiry into relevant themes as they arise in the course of the fieldwork.

The FDGs allow interaction with many people at the same time, increasing the reach of the evaluation. The discussions allow examination of different themes and receipt of consultative feedback from a range of actors. From this, views can be triangulated both within the group (from each FGD) and between the groups (based on other FGDs and interviews). The aim was for each focus group to involve 6 to 10 participants, though as discussed in Annex I, which provides details on the fieldwork, some amendments were made during data collection.

4.2.2.1   Respondents

The sampling of respondents and type of instrument used differed slightly from the BL, and is set out in Table 5 below.

Table 5:   Sampling and instruments by participant group

Participant	Sampling	Type of instrument	Change from baseline?
Head teacher	No sampling. In the head teacher's absence the assistant head teacher to be interviewed. The head teacher was interviewed in each school, though in one case the head teacher in place was temporarily assigned to the position, awaiting formal appointment.	KII	Same
Community Leader	The village committee chairperson.	KII	Same
Teachers	Maths and Kiswahili teachers teaching standard 1-3, including only those teachers who had received EQUIP-T INSET away from school. If there are more than 8 such teachers, 8 of them will be selected randomly to participate. If there are fewer than 4 teachers, we will consider carrying out two KIIs instead of an FGD.	FGD	Same, although last time at BL no teachers had received EQUIP-T INSET and teachers from higher grades were invited if there were not enough teachers in standards 1-3.
School Committee	All members of the SC were invited, aiming for attendance of 4-10.	FGD	Same
Parents – Fathers and Mothers separately	10 fathers and 10 mothers (not from the same family) were selected at random from a list of parents of children in standard 1-3. The list of children's names were selected at random the day before arriving (using the register collected by the quantitative team), and the HT was asked to call those children's parents to ask them to come to school the following day. The randomisation is expected to produce a group with some heterogeneity around socioeconomic status and religion.	FGD x 2	Change: For BL fathers and mothers were interviewed together. For midline they were interviewed separately to give each group a better chance to speak freely, particularly on issues that might be gender-specific. At BL the HT selected the parents based on guidelines, this time randomisation was used in the first instance to avoid any bias.
Children	3 boys and 3 girls randomly selected from standard 3.	FGD	Same
Ward Education Coordinator	No sampling – relevant WEC for the school	KII	Same
District Education Officer	No sampling – relevant DEO for the school	KII	Same
Regional Education Officer	No sampling – relevant REO for the school/district	KII	Same
Regional Team Leader (EQUIP-T Managing Agent)	No sampling – relevant RTL for the school/district	KII	Same
National EQUIP-T Managing Agent staff	Component technical leads and National Coordinator	KII	Same
Source: OPM IE team.			

4.2.2.2 Development of tools

The tools were designed in a comprehensive, integrated manner to ensure that each theme is assessed not only through the perspective of the people immediately related to that theme, but also other respondents who are likely to know about those issues. For instance, when evaluating the effect of teacher training, in addition to teachers and head teachers, the study also asked parents, community leaders, children, and local and district level education officers about the changes in teaching and possible contributing factors. Such a composite approach will ensure the rigour of the quality of data through triangulation.

The tools were largely developed based on the evaluation matrix for the ML (Volume I, Annex B) which was used to focus the questions for the qualitative work. This process was carried out in tandem with the revisions of the quantitative survey instruments, to ensure complementarity across the two methods as far as possible, and to use the qualitative visits to follow up on findings from the quantitative survey. The BL research guides were also used to inform the tools. Generally, the ML tools contain more specific topics and less exploratory discussion than in the BL. This is because the tools are now looking to see how things have changed and the extent to which EQUIP-T interventions or other factors are accredited with impacting change. In comparison, the BL was used more for understanding the context, issues, and relevance of the programme intervention design.

Qualitative tools were tested during a one-day pilot on the fourth day of training. The pilot particularly focused on identifying concepts and questions that may be misinterpreted by respondents, or may come across as leading. It was also important to identify programme specific terms such as 'PTP', to ensure that the tools were able to gather data around these concepts. The final day of training was then spent on refining tools based on pilot experiences, as well as additional training of researchers based on their experiences in the pilot and the emerging findings.

4.3   Fieldwork

4.3.1   Overview

The fieldwork was split into two parts, with one team conducting the school-based research and another carrying out the KIIs with education managers and officials (at ward, district, regional and national levels). Training of the school-based team took place in April 2016 in Dodoma. This comprised classroom-based and field-based training in schools with similar characteristics to those in the sample. Following training, the school-based team spent three weeks (April/May 2016) collecting data, spending one week per district (one day per school, leaving two days per week for debriefing and transcription). The qualitative team visited schools approximately one week after the quantitative survey team visited.

The second team, comprising one lead researcher and a translator, spent two days at national level interviewing EQUIP-T staff, followed by 10 days interviewing education official and managers across the three districts. The timing of this research overlapped with the school-level research enabling the two teams to discuss emerging findings during the fieldwork. For more details on the fieldwork, including the composition of the teams, key challenges in the data collection, as well as the quality assurance processes followed to ensure rigour, see Annex I.

4.4 Analysis

The analytical approach to qualitative data uses applied thematic analysis, primarily to confirm a set of hypotheses also known as classic content analysis (Krippendorf 2004). The selected principal approach, 'confirmatory analysis', aims to confirm a set of pre-existing hypotheses and generates codes from these hypotheses that are applied to the data. This is in contrast to exploratory analysis that derives hypotheses from the data collected (Guest et al. 2012). Exploratory analysis was used as a secondary analytical technique, to ensure that the qualitative component is responsive to unexpected information and that all relevant themes and hypotheses around EQUIP-T have been identified.

Applied thematic analysis requires the researchers to interpret data collected, i.e. the textual record of the transcribed and translated interviews and focus group discussions (Guest et al. 2012). It does not rely on counting words or phrases, but identifies and describes implicit and explicit ideas that are organised into themes.

Initial brush-coding[6] was done (using Nvivo 11[7]), in which a sample of transcripts from one school were used to identify codes based on what was inductively emerging from the data. Thereafter the set of hypotheses in the qualitative evaluation matrix provided a set of themes to which all the data was coded. The combination of inductive initial brush-coding and coding based on the evaluation matrix was done to:

- mitigate bias of analysis (where a researcher would only pick up on findings directly linked to the research questions), and
- ensure an analysis close to the midline evaluation matrix, whilst continuously allowing for the emersion of nodes[8] not directly in the matrix, in line with exploratory analysis.

The strength of each piece of data was considered in the light of the context it came from (for instance the knowledge that the person cited is likely to have about the subject, the incentives they may have to respond in particular ways, and the corroboration from other qualitative sources). Researchers then assessed the balance of these groups and whether the conclusions support the initial hypotheses. This is based on both the frequency of responses (without claiming to be statistically representative) and the comparison between the views expressed. This analysis thus enables triangulation both between participants and schools, and thematically within each transcript itself.

As case study schools were sampled, researchers assessed the difference between case study schools to be able to consider the situational reality of findings, however, thematic findings did not differ between case study schools at ML and as such a broader thematic analysis has been emphasised in the report. Likewise, findings were considered in light of both who the respondent was, as well as the various incentives a respondent might have to share certain views (discussed throughout the report in terms of social desirability bias).

As ML focused on reasons for change, and questions were less exploratory but tied directly to the programme, researchers found that thematic findings were strongly consistent across respondents. In cases where views differed, these were discussed and analysed to assess the possible explanations for why views differed. Additionally, where some respondents (such as parents) were

---

[6] School 1, District A transcripts were analysed independently from the evaluation matrix in order to ensure that codes emerging were based in the data and to avoid bias based on what information the researcher expected to find. A coding structure was developed based on this, which was then checked against the evaluation matrix.

[7] Qualitative analysis computer software.

[8] Nodes are 'codes' (thematic groupings of data) in Nvivo 11.

unable to comment on certain questions (such as the Parent-Teacher Partnership (PTP)), the analysis considered why that might be, and thus included 'gaps in evidence' as evidence in itself. For readability, the report thus chose to refer to 'case study schools' or 'respondents' in cases where consensus was strong across participants groups (including head teachers, teachers, SCs, CLs, fathers and mothers). Children were not consulted on all topics, and their views are thus included in the analysis around those topics the FGD touched upon – in particular for Component 1 (teachers) and Component 4 (community).The analysis was conducted by researchers who also conducted fieldwork, thereby helping to ensure that errors of interpretation were minimised.

The depth of interpretation and analysis at ML is affected by the wide scope of the qualitative component. As the EQUIP-T TOC covers four large components, the evaluation matrix is likewise very extensive. In order to explore questions around all components, there was less space for researchers to probe in-depth around each answer and theme – as FGDs and KIIs cannot exceed certain time-limits to avoid both participant- and researcher fatigue. This means that the below analysis, whilst balancing description and interpretation, tends to lean towards being descriptive.

The research team has been careful when assessing evidence beyond description where evidence has not been sufficiently thick, in order to avoid faulty interpretation of the meanings and reasoning behind answers. Still, this is not to say that a descriptive account is not valuable. The qualitative data has been able to capture a range of individual responses around each area of the EQUIP-T TOC, and has seen great consensus amongst respondents on various issues and themes. As such, though qualitative findings never aim to be representative, the team is confident that the qualitative analysis can be used as indications of larger themes for the programme to consider. The breath of findings of the qualitative data is a key strength of the below discussions, and allow for themes (and discussions around Components) to be considered from a variety of angles.

Additionally, as the qualitative team aimed to be reflexive in data collection and analysis, allowing for new themes to emerge, findings resembling BL analysis are reported. Endline will allow for further assessment of these new themes. This also accounts for findings around particular activities such as PTPs. PTPs in many schools, whilst elected, have not been active and as such findings are prior to PTP activity. Lastly, as head teachers in many schools were new, the qualitative analysis around Component 2: SLM, is largely in line with a BL analysis.

The interviews with senior managers were analysed separately to the school-level data, as the focus of these two data sources differed. The team leader similarly conducted thematic analysis (using Excel), noting where transcripts referred to major themes and the detail of arguments, and pulling out key quotations. The transcripts were continuously re-visited to check the weight of the arguments and the context of the interview that may have impacted respondents' responses. The findings from the district, regional and EQUIP-T HQ interviews where then discussed with the research team focusing on the school-level data in order to cross-check analysis, discuss findings and triangulate across research streams.

4.4.1 Addressing rigour through the analysis

Rigour in the analysis of the qualitative data comes from six principal sources:

- First, through rigorous training of field researchers prior to field work.
- Second, extensive quality assurance procedures implemented during data collection (see Annex I).

- Third, by including different members of the team in analysis (at debriefs as well as during coding, analysis and write-up), in order to ensure reliability and consistency in analysis, as well as to mitigate single-researcher bias.

- Fourth, the analysis sheet (including de-identified data and the applied data codes) is available for external scrutiny, with confidentiality controls.

- Fifth, through the comparison of different data sources, both qualitative and quantitative. This was done initially during a workshop held in early August 2016 with the quantitative and qualitative work streams in order to discuss findings and help identify strengths, gaps and areas for further exploration in the data. Further information sharing and enrichment of the analysis continued into the report writing phase (see Chapter 2 for more details on mixing methods at analysis stage).

- Finally, the analysis was subjected to internal and external peer review (see Annex B).

4.5   Risks and limitations

The main limitations of the qualitative component are presented in Table 6. Column one states possible limitations whilst column two explains why and how these can be mitigated. In addition to these general limitations, a brief description of some specific problems and issues that were encountered in the ML qualitative data collection and analysis is given below the table.

Table 6:   Possible limitations of qualitative component

Possible limitation	Why this is limiting and mitigating factors
Inference beyond the selected research sites is limited.	While the qualitative data of EQUIP-T examines perceptions at multiple levels of the education system, the findings of the research reflect the particular districts and schools selected. This can be mitigated to some extent by purposively selecting the research sites to have as much potential for generalisation as possible, however there remains a risk that the findings are affected by the choice of districts and schools.
Given the non-representative nature of the qualitative selection of districts and schools the information provided will be indicative	The qualitative component of the IE offers nuanced first-person accounts of people's perspectives and experiences without claiming that these accounts are representative of other similar communities' and schools' experiences. When considered together with the representative quantitative results, the qualitative findings provide perspectives on underlying issues including potential explanations for results identified in the quantitative evaluation and of factors that can determine the success of a programme such as EQUIP-T.
Participants in the qualitative study are likely subject to social desirability bias in their responses	It is common for participants to respond to the facilitator's questions with answers they think the study want to hear. This becomes particularly prominent at ML and EL as participants know the research is about EQUIP-T and are more likely to speak positively. To mitigate this response bias, facilitators were trained to carefully probe further around responses, and give opportunities for respondents to mention other factors that would support or contradict the positive accounts. Additionally, social desirability bias was mitigated at the stage of analysis, where researchers carefully explored and discussed the data to ensure the strength of findings. This included the analysis of social desirability bias in itself, and how it may be a useful indicator of changes in awareness, whilst not necessarily indicating a change in attitudes or behaviour. .

The qualitative part of the impact evaluation covers all four programme components and is thus very large in scope which leads to a vast amount of data collected from each case, rather than a focus on depth.	The breadth of the EQUIP-T components being implemented at multiple levels necessarily puts constraints on the ability of the qualitative research to analyse the impact of each component in depth. Nevertheless, the qualitative data generated during the KIIs and FGDs, taken together with the quantitative findings, offers a basis from which to draw conclusions about areas of strength and weakness in the EQUIP-T programme.
Inaccurate identification of EQUIP-T interventions by respondents	The various EQUIP-T interventions as well as other programmes were at times confused by respondents, and different acronyms in Kiswahili adopted in different areas. This may result in respondents inaccurately identifying EQUIP-T interventions, or for data to be missed due to language and translation discrepancies. In order to mitigate this risk, training was centred around each researcher fully understanding each intervention rather than simply asking questions on a concept. Instruments were designed to initially explore whether respondents were aware of the named interventions, but then the topic was broadly discussed and questions asked around it. For example respondents were asked whether there was 'any group in the school that parents took part in' in order to capture understanding beyond the official terminology for PTPs. Additionally, as terms were identified during pilot (for example UWW and Kikasi Kezi) these were used in probing, but the holistic approach to questioning around each intervention meant that an answer around terminology was not considered definite. Lastly, at analysis stage thick evidence was prioritised, for example, though respondents may say that had not heard of a PTP, they were aware of having elected parents to be part of a committee. Analysis was thus based in a solid understanding of interventions, and evidence analysed in light of what respondents knew regardless of whether they were aware of it being EQUIP-T specifically.

The ML qualitative data collection and analysis encountered a few specific challenges:

- First, the significant head teacher turnover in case study schools meant that an important key informant could not always be considered a key informant, but the interview rather resembled a standard semi-structured interview. Key informant interviews are conducted with those individuals who are likely to be deeply knowledgeable of the whole case sampled, as well as on specific issues explored in the research. When head teachers are new, they are thus less able to comment on changes in SLM, as well as have less experience of the case school in which they are now working. Additionally, as several head teachers had not attended SLM in-service training (INSET) the qualitative study could not fully conduct a ML analysis of this component. Yet, as KIIs with head teachers were conducted by a highly experienced team lead (together with a translator), the interviewer was able to restructure interview instruments to allow for exploration of themes more in-line with the experience of each specific head teacher. This meant that though head teachers might not be able to answer specific questions around EQUIP-T, the research gathered a wide breath of information around themes through these interviews.

- Second, PTP members were not explicitly sampled as part of the parent sample, because of the broad range of research questions to be covered, the need for perspectives from different socioeconomic, religious and gender groups, and resource constraints (which meant that sampling trade-offs had to be made). By chance some PTP members were part of the parents' sample, and information on PTPs was also gathered from other respondents (including HTs, community leaders, teachers and the SC). There is a risk therefore that the

analysis of PTP activities and functioning misses some themes, as it is not able to rely on sufficiently 'thick data'. The analysis of PTPs was highly reliant on FGDs with teachers, as they would have been aware of activities and functioning of PTPs. That parents interviewed were often not aware of the PTP's existence was in itself an interesting finding around the wider spill-over effects of parental involvement in schools. Indeed, the PTP is supposed to link all parents closer to the school/classroom, so if it is functioning effectively then the wider parent body, not just the 14 representatives, should know about it and its activities. There is a risk that if the parent sample was purposively targeted at PTP representatives then this may yield biased responses about PTPs' functioning, and induce bias more generally if these people are/were already more proactive in school life.

- Third, as discussed in the section on analysis, the qualitative part of the IE covers all four programme components and is thus very large in scope. This leads to a vast amount of data collected from each case rather than a focus on depth. During data collection this meant that instruments were lengthy, and the team faced challenges covering all questions with all participants – placing limits on the degree of probing that was possible. The research was thus able to gather data around each component with each group of participants, whilst not always gaining sufficiently 'thick' data for deeper analysis and interpretation. Yet, the breadth of data at ML provides a strong basis for endline design, where themes and issues assessed to hold particular significance for the programme can be focused on and explored in-depth. As ML was able to gather such extensive evidence around each component, there is scope to narrow the qualitative part of the evaluation at endline to conduct a targeted in-depth exploration of specific themes complementary to the quantitative IE. There would be considerable advantage in narrowing the focus of the endline qualitative research. As such, the potential trade-off between depth and breadth will be carefully considered and discussed with DFID and other stakeholders during the endline design phase.

5 Cost study design and ML methods

This chapter contains background and methodology on the costing study to supplement the main analysis chapter included in Volume I.

5.1 Objectives of the costing study

The impact evaluation includes a costing study, intended to understand the costs of the programme's components and therefore the cost and affordability of the programme being continued after the DFID funding has finished. The objectives of the costing study are to:

1. review and present programme expenditure, if possible by sub-component/activity, by region, and over time, in order to compare different levels of expenditure;
2. estimate unit costs of activities, for example of delivering literacy INSET to one teacher;
3. estimate the costs of expansion of the programme, in order to understand what costs would need to be met if (elements of) the programme was taken up, continued and further rolled out by the government;
4. put these costs within the fiscal context – in terms of both the current budgets and levels of spending of districts (LGAs) and government as a whole on education, and in terms of the projected macroeconomic landscape.
5. In addition, with the costs set out, it is possible to give information on the cost-effectiveness of the programme, i.e. the cost of generating a change in pupil learning outcomes, which is useful for comparison of models and intervention options for targeting quality education.

The ability to conduct detailed, nuanced analysis depends on the availability of detailed and consistent financial expenditure data from EQUIP, obtaining sufficiently detailed implementation data, and accessing data on wider government spending on education.

5.2 Analysis at midline

5.2.1 EQUIP MA PSA spending

Up until November 2015, all funds spent on the EQUIP programme activities were managed entirely by the managing agent (MA). From November 2015 onwards, some spending continued to be managed by the MA whilst there was also spending being carried out by LGAs through the government's own systems.

The EQUIP-T MA supported the costing study by sharing its expenditure data. The budget tracker, showing all spending up to June 2016, is comprehensive of the implementation of EQUIP-T so far, but limited in its detail and consistency, and therefore the level of analysis it allows. The data contains detail which allows presentation of spending over-time and by sub-component (there are five components, and a total of 33 sub-components). The sub-component definitions vary in terms of detail, and so, for example, it does not identify specifically the spending on INSET for teachers. It is also not possible to disaggregate MA spending by region or district.

The MA also supported OPM in extracting a further level of subcomponent detail from the accounting software. This pulls out each item of spending, for example under subcomponent 1.2, there were over 900 entries. At this level, the descriptions are still relatively ambiguous (e.g. 'payment for workshop') to be able to identify exactly what the activity was. In addition items have been coded under categories that should now be redundant: activities that were planned at the programme's inception (such as INSET for Standards 4 to 7) but have since been replaced by

other priorities. The details of a workshop content – such as a module development workshop, or an orientation workshop, or a detailed training on modules 1-4 – could be found in supporting documents with payment vouchers, in hard copy. It is not feasible to look at this detail for so many entries.[9]

As such, it is not possible to classify costs according to whether it was fixed/one-off (such as materials development), semi-fixed (such as training regional management on aspects of the programme) or variable (such as training on school leadership and management, which may be repeated at various frequencies, or distribution of textbooks which would need replacing). It is also not sensible to make assumptions about types of costs over time (such as assuming a first phase of fixed cost, a second of semi-fixed and a third of variable costs. This is because the programme keeps evolving and developing new activities – such as the maths INSET, the school clubs, and the school information system roll-out. There have been development activities throughout the period of implementation. Without ability to classify these different types of expenditures, accurate unit costs cannot be applied to project roll-out or expansion of the programme.

As an alternative, crude unit costs have been estimated, using broad categories of spending and high level monitoring data. For example, the average spending on improving teaching (Component 1, excluding the costs of the School Readiness Programme) per school has been calculated. For these unit costs, the units were taken from EQUIP-T's monitoring data, as at March 2016. The unit costs are therefore 'crude' for two reasons:

- The cost includes everything spent so far on that activity, so does not present the marginal cost of expanding the programme to include one more unit.
- There are many potential interpretations of units. For example, the number of beneficiaries of early grade INSET could be different for each set of modules delivered, and for the 3Rs curriculum training. Likewise, the number of schools can change as schools open or close. However, one measurement of the unit is chosen.

To support interpretation of the figures, different types of units have been calculated – such as spending per school, spending per trainee, or spending per pupil. The monitoring data includes overall primary enrolment rather than specifically enrolment in Standards 1 and 2.

5.2.2 LGA PSA spending

From late 2015, the funding modality changed such that a substantial portion of programme funds is transferred directly to LGAs, rather than through the EQUIP-T programme. LGAs received their first tranche of funds in late November 2015, with the exception of Mara which received their first tranche in February 2016. A second tranche was received in early June 2016. Spending is budgeted and ring-fenced into ten codes in the government financial management system called Epicor. Districts report their spending through Epicor, and send monthly expenditure reports to the Regional Administrative Secretariats, who forward them to EQUIP-T.

The MA also supported the analysis of LGA spending, providing LGA spending consolidated reports, individual reports for three LGAs to focus on INSET, and supporting with explanations. Fund officers at the MA work hard to minimise any errors made by LGAs in coding of spending.

Unit costs have been calculated at the regional level based on the regions' total spending and monitoring data held at EQUIP-T headquarters. The HQ monitoring data may not accurately reflect

---

[9] It should also be noted that when a small sample of entries was examined in detail, mis-classifications were found – such as spending for the SRP which came under the wrong subcomponent.

the actual units benefiting from decentralised LGA spending. For example, the MA does not have aggregated data on the number of beneficiaries of INSET since decentralisation, and instead the number of beneficiaries from a previous module training is used. The regional unit cost analysis presented here for midline would be difficult to update at endline, as the SRP costs will come under the same code as 3Rs INSET.

5.2.3   Early Grade INSET model cost analysis

In order to assess the feasibility of rolling out the EQUIP-T training model further, the spending of LGAs on delivering EG INSET was examined. Through discussion with fund officers in EQUIP-T MA, three LGAs were selected to give a range of examples of training costs:

- Kilwa District Council in Lindi – to give a rural LGA;
- Bariadi Town Council in Simiyu – an urban LGA that adapted the delivery model to use more cluster-based training;
- Tabora Municipal Council in Tabora – to give an urban LGA.

The monthly spending reports from these LGAs were used to identify the EG INSET activities provided between December 2015 and May 2016. The numbers and types of participants, type and length of training, and module topics were found in the reports. From this, a unit cost was calculated to show the average cost per beneficiary of the whole training, and on a per day basis. For the sake of simplicity, WECs, Quality Assurers and District INSET Coordinators and school-based teachers were included as beneficiaries, District staff such as DEOs, cashiers, drivers or security staff were not included as beneficiaries. The more complete description of this training is given in a table in Annex J and it should be noted this is only the cost of the training itself, and does not include the printed modules given to participants.

5.3   Analysis at endline

For the endline study, high level spending and unit cost analysis can be updated to include a further two years of data. It will still not be possible to separate out specific activities to estimate refined costs of expanding or continuing elements of the programme, but cruder, bulk replication of the project could be estimated. At endline, the cost analysis can be put in the context of recent government spending on education from its own budget. This will require access to data on the budgets of MOEST, PO-RALG and the regions and LGAs on education. Similarly, macroeconomic trends and forecasts can be used to get a sense of feasible budget allocations to education, to put in context the affordability of the EQUIP-T model. In addition, using final estimates of impact on learning outcomes, we will be able to show the cost of changing learning outcomes by a standard unit, for comparison of EQUIP-T with other education improvement programmes.

5.4   Limitations of the cost study

There are a number of limitations to this analysis, largely relating to the availability of data, many of which have been outlined above. The key limitations are as follows:

- There may be some miscoding in the PSA data from EQUIP-T MA, and to some extent from the LGAs too. Whilst this is not expected to be substantial, it does mean that spending on some activities may be over or underestimated, and this is more likely the more specific the activity and analysis becomes.

- The level of detail in the spending classification is not sufficient to classify the costs into one-off/fixed versus variable. Thus it is not possible to fully estimate the costs of rolling out or continuing parts of the programme.

- The analysis on the comparative unit costs of different INSET models does not give any indication of the quality of the different models. Evidence on the effectiveness of the different approaches would be needed to make a comprehensive decision on optimal INSET models.

- This analysis only includes the expenditure categorised as programme support activities. By excluding technical assistance, the analysis is underestimating the full cost of developing and managing the materials and programme. Thus, if the government were to continue the programme, this analysis may underestimate the costs if additional management support were required.

- Furthermore, the analysis does not include the opportunity cost, or other costs, borne by public officials in taking part in the programme. For example, the time spent attending INSET which could otherwise have been spent on lesson preparation or teaching. This analysis therefore does not present the full economic cost.

- At endline, the comparison of costs with the change in learning outcomes (for the purpose of cost-effectiveness analysis) will be affected by the challenge of attribution. In particular, whilst the impact on learning can be viewed as the 'impact of EQUIP-T over and above implementation of the LANES programme,' the costs may not be additional to LANES in the same way.

Part F Evidence

6 Impact estimation

This Chapter explains the measurement approach taken to impact estimation, and presents the detailed results.

6.1 Impact identification strategy

A rigorous identification of programme impact in quantitative studies generally builds on the idea that such impact can be defined as the difference in the outcomes measured among individuals that participate in a programme compared to the outcomes measured among the same individuals in a theoretical state of the world where the programme is not implemented but where everything else, except the programme, stays the same. This is normally referred to as the counterfactual and, because it is purely hypothetical, they key challenge that impact evaluations face is to find alternative observed counterfactual measures that can credibly be used to infer programme impact.

A Randomised Controlled Trial (RCT), where observations are randomly assigned to a treatment and control group, is commonly considered as one of the most robust designs to deal with the problem of the counterfactual. Because treatment assignment is implemented randomly in these trials, individuals from control and treatment groups are, on average, the same. This means that after the implementation of the programme, averages of outcomes measured among participants and non-participants can be compared directly and differences can be attributed to the programme, rather than any other confounding factors. Sometimes, however, implementing an RCT is not feasible or not appropriate. Alternative identification strategies use econometric modelling techniques to try to come as close as possible to replicating the situation of such an experimental design.

This was the case in the present evaluation, where an RCT was not feasible and schools were assigned to participate in the programme based on programme management decisions and some pre-defined characteristics. Control schools were selected to match those characteristics.[10] Specifically, the quantitative impact identification methodology used in this study follows a quasi-experimental design that combines two approaches: Propensity Score Matching (PSM) and Difference in Difference (DID) analysis. This allows to combine the strengths of both of these methods in order to robustly estimate the difference in key impact indicators across treatment and control schools that can be attributed with statistical confidence to EQUIP-T. The following sections elucidate how both PSM and DID were implemented and combined in the current evaluation.

6.2 Propensity Score Matching

The key problem that PSM attempts to solve is selection bias. In the present case, this problem appears because pupils and teachers from schools that did receive EQUIP-T support could be systematically different from individuals in schools that did not receive such support and form part of the comparison group – because the assignment to treatment status was not implemented randomly. Such systematic differences could plausibly be related to outcome measures that this evaluation is interested in. This in turn implies that observed dissimilarities in outcome measures across individuals from treatment and control schools could be due to underlying systematic differences and not the programme itself. Simple comparisons of indicators across such groups

---

[10] Note that the term 'control group' is used throughout this document to refer to the comparison group.

would be invalid and biased to infer programme impact, because these groups cannot be assumed to be alike. This is the problem of selection bias.

PSM tackles this problem by using data from the control group to construct appropriate comparisons to pupils or teachers in the treatment group, thus building a valid counterfactual. This happens by matching and comparing outcomes for units in the treatment group with control units that are as similar as possible to each other according to a set of relevant observable characteristics, i.e. comparing like with like only. Relevant characteristics are the ones that are thought to be driving selection bias. These are the characteristics that are systematically different across treatment groups and are related to outcome measures of interest. When appropriately controlling for all of those characteristics, selection bias is also controlled for. A good example is represented by the age of the head teacher. In the present analysis, this is found to be significantly correlated with both pupil Kiswahili and Mathematics outcome measures as well as teacher absenteeism rates. This could plausibly indicate that the age of the head teacher is a proxy for experience that positively influences learning and teacher level outcomes at schools. It is therefore a 'relevant' characteristic to control for in the PSM model. While the variable mean value is unbalanced between treatment and control groups before matching, the model achieves balance on this after the PSM matching procedure. The matching therefore successfully controls for this sample characteristic, increasing the comparability between treatment and control groups and making the estimation of impact more robust.

Specifically, PSM is a two-stage analytical approach that employs a propensity score as a 'comparator metric' that summarises the information of the set of relevant characteristics, i.e. the ones that drive selection bias, defined above. This propensity score can also be interpreted as an estimation of the hypothetical probability of any individual to be in the treatment group, given its characteristics. The first stage of any PSM analysis is to compute a valid propensity score for each unit of observation. The second stage is to then compare outcome indicators of interest across units (i.e. teachers or pupils in this case) with similar propensity scores. Note that because outcome indicators from treatment units are compared to outcome indicators from specific control units based on the propensity score, the estimated average treatment effect will be valid for the group of treatment observations only. This means that PSM allows to estimate an Average Treatment Effect on the Treated, or ATT.

It is important to note that, for PSM to work appropriately, the comparator metric constructed in the first stage needs to be valid. For that to be the case, it needs to be calculated using variables that are not influenced by the intervention and are 'relevant' for the construction of the counterfactual. As described above, 'relevant' here means that these are variables which are driving selection bias. To meet the first condition, and given the need to match treatment and control groups at both BL and ML, the model constructs propensity scores only using variables that can objectively be considered as being predetermined and hence not having been influenced by the programme at the ML stage of the evaluation. In order to meet the second condition, researchers typically argue from a theoretical perspective about which variables could be relevant to control for selection bias. This study improves this selection of relevant variables by using a data-driven algorithmic approach that aims to reduce researcher discretion in the choice of variables. See Chapter 6 in Volume II for a list of variables included and for a detailed description of the variable selection algorithms employed here.

The validity of any PSM approach also depends on how well it reduces any imbalance, and thereby selection bias, between treatment and control groups. Achieving balance means that if matched appropriately treatment and control groups' characteristics will not be significantly different from each other. In other words, this means that, across the list of relevant characteristics that are assumed to drive selection bias, the treatment and control groups will be statistically

similar to each other. See Chapter 6 in Volume II for a detailed description of how covariate balance across treatment and control groups was assessed in the present case.

However, even after implementing a matching procedure, some imbalances across treatment and control groups can remain, which potentially could invalidate an impact identification strategy unless further analysis is implemented. This was the case for some pupil and teacher level indicators in the present evaluation – where imbalance remained across some characteristics of pupils and teachers after implementing PSM. As mentioned above, to address this remaining imbalance, this study combines PSM with a DID analysis.

PSM first stage model selection

To estimate the propensity score in the first stage, this study followed the procedure suggested by Imbens and Rubin (2015, p. 281 ff.). The underlying model specification for this procedure is either a logit or probit regression for the first stage. This means that the propensity scores are estimated by first specifying treatment and control assignment as a binary variable that has the values 0 (for control) and 1 (for treatment). The estimated scores are then modelled as the fitted values that are derived from a logit or probit estimation, with the binary treatment variables as dependent variable and the covariates across which balance is supposed to be achieved as the regressors. These fitted values lie between 0 and 1.

To be more concrete, in the case of a logistic regression specification, the binary response variable is modelled as follows:

$$\Pr(T = 1 \,|X_i) = \frac{e^{f(X_i)}}{1+e^{f(X_i)}}, (1)$$

where $\Pr(T = 1 \,|X_i)$ is the probability of the treatment indicator (T) being equal to one, conditional on the covariates (X_i) for unit i. The function $f(X)$ is normally modelled linearly, i.e. is of the form $f(X) = X\beta$. The coefficients of this function (β) are estimated using maximum likelihood techniques. The fitted values, i.e. the predicted probabilities that follow from this procedure, are the propensity scores for each unit of observation.

The key question for the first stage is which covariates to include in in $f(X)$ so that this procedure produces a valid estimate of the propensity score. Building on the procedure described in Imbens and Rubin (2015) for selecting covariates, this study implemented a four-step approach to make this decision:

1. Select a set of basic covariates based on substantive grounds

The starting point for the PSM analysis was to select variables that were likely to be relevant and valid to be used for this analysis from a theoretical perspective. 'Relevant' in this case meant that variables had to be selected that were theoretically expected to be correlated with treatment status and treatment effects, thereby introducing selection bias in a simple comparison of treatment outcomes between control and treatment groups. This requires a theoretically substantiated understanding of the relationships that were being analysed.

'Valid' in this case meant that variables had to be selected that were expected to not be influenced by the programme. This is because PSM analyses had to be implemented both using baseline and midline data. At baseline, this would not be an issue, given that the programme had not started yet and no variable would have been influenced by the programme. At midline, however, this was not the case. Hence, only variables were used for PSM for which a plausible argument could be made

that they have not been influenced by the programme at midline. The set of linear covariates that were fed into the first stage selection process for pupils and teachers is outlined Annex H.

2. Increase the set of valid covariates based on algorithmic approaches

In addition, this study employed variable selection algorithms to identify valid variables, i.e. variables that were not affected by the programme, and that are significantly correlated both with the treatment status and the outcome variable. There are a variety of methods available to do this. This study's approach was to implement stepwise regressions. Such regressions are commonly used and easily implemented algorithms to select independent variables based on significant correlations with certain dependent variables.

There are two stepwise regression approaches that can be employed for this: backward and forward stepwise regression. The underlying idea behind both approaches is to check each covariate, step-by-step, for significant correlation with the outcome and treatment assignment variable separately. Such a correlation is relevant because variables that possibly bias impact estimates will have some relation to both the treatment status and the outcome looked at.

Backward selection starts with the full set of covariates, i.e. a regression including all variables, and then discards the term that is least significantly correlated with the dependent variable. It continues to do so until all variables that are uncorrelated with the dependent variable are discarded. Forward selection, instead, starts with an empty set of covariates, i.e. a regression on a constant, and then checks the significance of each covariate if it is included in the regression. It then adds the most significantly correlated variable to the model. This step is repeated until all significant covariates are included in the model.

Both for backward and forward estimation a threshold p-value for what is considered to be significant needs to be specified. For backward selection, this means setting the level for identifying whether all variables that are uncorrelated with the outcome variable have been discarded: if the p-value of the least significant variable remaining is under the threshold, i.e. all the variables still included in the model are even more significant, the procedure stops. For forward selection, this means setting the level for identifying whether all significant covariates have been included in the model: if the p-value of the most significant variable to be added is equal to the threshold, i.e. the significance level of all variables that have not yet been included in the model is equal or below the threshold, the procedure stops. Setting this threshold therefore influences the variables that are selected in stepwise regressions.

This study implemented both backward and forward selection, using baseline data and using thresholds of $p = 0.05$. The analysis is employing this covariate selection procedure on both relevant outcome variables and treatment status, given the importance of determining the significance of covariate correlation on both, as explained when discussing our approach above. A common set of variables for the models were then selected based on whether they were selected in either of the forward or backward stepwise regressions.

3. Increasing the set of covariates with polynomial and interaction terms using algorithmic selection

In a third step, the same method of stepwise regressions (backwards and forwards) was employed to augment the set of covariates by quadratic terms or interactions of variables that had already been selected in steps one and two. The rationale behind this is the fact that balance might only be achieved if the propensity score is estimated using non-linear transformations of the variables selected in the first two steps (Imbens and Rubin 2015, p. 287). Again, the stepwise regression approach helped to decide which of these non-linear terms were significant predictors of differences across control and treatment groups, and should therefore be controlled for.

4. Assessing whether the set of covariates (including polynomials and interaction terms) could be used at midline as well

The full variable selection process described above was implemented using baseline data first. In a second step, this model was transferred to the midline datasets and – using the variables selected at baseline – dummy regressions were implemented to assess whether these models could be used at midline as well. Problems could appear where collinearity between some covariates would appear at midline that could potentially make propensity score estimation unstable. When this happened, these covariates were removed from the estimation process.

The result of this process was the identification of an optimal selection model comprising a set of covariates that were included in the first stage estimation of the propensity score. This four-step approach was conducted for every estimation strategy for each of the outcome variables. It is important to note, however, that good balancing properties using PSM also depend on the matching algorithm used in the second stage of the PSM analysis described in the next section.

It is also important to emphasise again that, as a result of the above process and in order to ensure consistency in the way the counterfactual was constructed and the comparison group identified at baseline and midline, the same set of variables was selected to construct propensity score at the two points in time, i.e. at baseline and at midline. This increases the degree of inter-temporal comparability between the two matched samples and improves the robustness of the estimates. See further below for an explanation of how baseline and midline estimates were compared across time.

Second stage algorithm selection

There are a variety of algorithms available to implement the second stage of PSM, i.e. to match control and treatment units to each other based on the propensity score estimated in the first stage. Figure 2 shows algorithm options and sub-options for each of these possibilities. It is beyond the scope of this report to explain in detail the technicalities of each of these approaches.[11] For all approaches the goal is to find appropriate, i.e. sufficiently similar, control group members for treatment group members. Differences between these approaches can be defined along three main dimensions: first, which estimated propensity scores are considered to be valid for inclusion in the analysis? Second, what is the appropriate range of propensity scores that define control comparators for treatment units? Finally, how are these comparators used when estimating the treatment effects?

The first dimension relates to the fact that within both control and treatment groups there could be estimated propensity scores that lie either at the upper or lower bound of the distribution, i.e. close to 0 or 1. For such values, there might not be an appropriately similar propensity score in the respective comparison group. However, for matching to work appropriately, there must be comparable propensity scores in both control and treatment groups – the so-called common support condition. Hence, matching algorithms employ cut-offs or trimming procedures by which some proportion of observations with propensity scores that are not comparable are dropped from the analysis.

---

[11] See Caliendo and Kopeinig (2005) for a summary overview.

Figure 2:   Matching algorithms selection
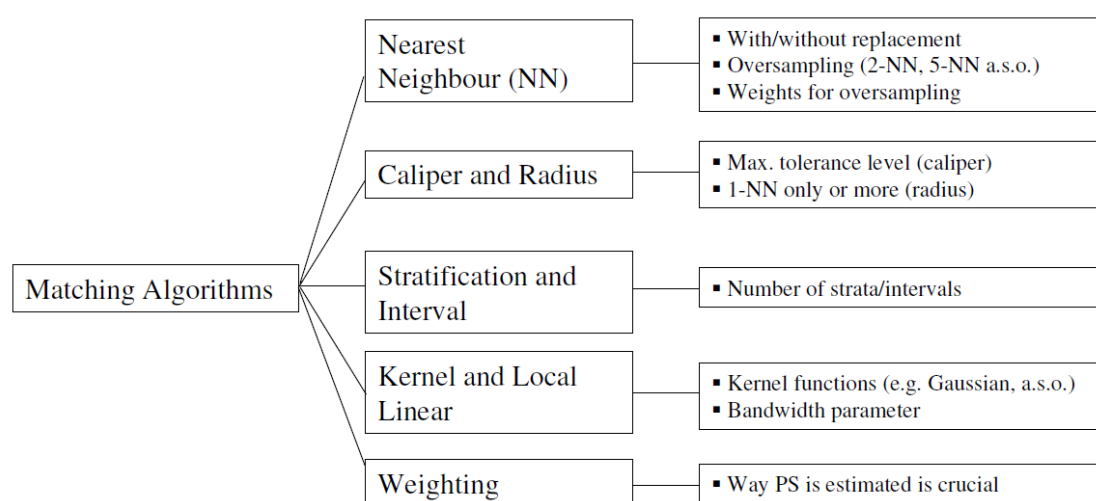
NN: Nearest Neighbour, PS: Propensity Score

Note: Figure taken from Caliendo and Kopeinig (2005, p. 9).

The second dimension relates to how units in the control group with propensity scores close to a treatment group observation are treated. For instance, kernel matching, as used in the main impact estimation for the PSM model, is a non-parametric matching estimator that uses the weighted averages of all individuals in the control groups to create the counterfactual outcome. The weights are determined by the distance between each individual from the control group and the participant observation for which the counterfactual is estimated. Therefore, higher weights are given to persons closer in terms of the propensity score of a treated individual (Caliendo and Kopeinig (2005), p.10–11). Alternatively, Nearest Neighbour (NN) matching with just one unit looks for the one control observation that has the closest propensity score to a treatment unit and compares the outcome measure for those observations. NN matching with more than one neighbour looks for several control units with similar propensity scores and compares the treatment outcome to an average of these neighbours. Caliper matching is similar to NN matching but does not include a fixed number of neighbours. Instead, the comparators are selected based on a maximum difference in propensity scores allowed.

Finally, the third dimension refers to how, once comparator units are found, the outcome measures are compared across treatment and control. For example, with NN matching and more than one neighbour simple averages are calculated. Similarly, with kernel functions a form of weighted averages are calculated to estimate treatment effects.

Selecting the appropriate matching algorithm for a PSM exercise is not straightforward and requires careful analysis of how well-balanced samples are after employing algorithms with certain sub-specifications. In general, however, the selection of models in this study was based on the fact that discriminating between models poses a bias/variance trade-off in the estimated treatment effect. For instance, in the extreme case of NN matching with just one neighbour, it could be that the NN is actually quite far away in terms of propensity scores and hence a bad match. If this happens often, this could introduce bias into the estimation procedure. A solution to this could be to implement matching using several comparators in a caliper matching setting. However, this could decrease the number of available matches, which could increase the variance of the treatment estimate.

Kernel matching with appropriate trimming and enforcement of common support is a good compromise between these different approaches and was therefore selected as the main matching algorithm for both baseline and midline estimates.[12] In order to find the optimal estimation model this study used different kernel matching algorithms with different bandwidths and trimming levels. These different results were then compared with respect to the best balancing properties, with the best performing approach being selected as the optimal. This was again conducted for each estimation strategy for each of the outcome variables and for both baseline and midline estimations separately.

Key PSM assumptions: common support and conditional independence

There are two key assumptions that need to hold for PSM to be a valid approach to estimating treatment effects: the common support assumption and the conditional independence assumption.

The **common support assumption** states that the estimated propensity score for all individuals in the treatment and control groups must lie within 0 and 1. Expressed differently, individuals in both groups must have a positive non-zero probability of belonging to either the treatment or control group and the distribution of those probabilities across the two groups must be such that comparable individuals across the groups can be found. This can easily be enforced by only comparing observations with appropriate propensity scores.

The second key assumption is the **conditional independence assumption**, which posits that, once observable characteristics have been accounted for, the outcome measure is not related to the treatment status anymore, other than via the effect of the programme. In essence, this assumption states that once observable characteristics are appropriately controlled for, treatment status can be treated as if it was assigned randomly. As described above, PSM deals with this problem by comparing outcome measures across treatment and control groups only for individuals that are similar, i.e. by controlling for the important characteristics that are related to both treatment status and the outcome measure. The conditional independence assumption simply states that all important characteristics have been taken care of. This means that any bias that arises due to participation in the programme has been dealt with. Note that this includes biases that arise due to unobservable factors – PSM cannot control for these and the assumption is that once observable characteristics have been dealt with no unobservable bias remains.

The validity of any PSM approach therefore crucially depends on how well the approach reduces any imbalance between treatment and control groups. Under conditional independence – i.e. independence of the treatment assignment from outcome measures when controlling for covariates – the propensity score is a valid balancing score. Conditioning on this score appropriately means that bias will be removed between control and treatment groups. Hence, treatment and control groups will be balanced, i.e. they will have similar covariate distributions. This means that, across a variety of different characteristics, the treatment and control groups will be similar to each other.

Assessing balance of covariates after matching is therefore a key step for any PSM analysis. The more balanced samples are after matching, the more plausible is it that the conditional independence assumption holds. As described above, however, balance also depends on the models and algorithms used to implement matching. The following paragraphs explain in detail how balance assessments were implemented and used in the current study.

---

[12] See Caliendo and Kopeinig (2005, p. 10 f.) for a short summary of the pros and cons of different matching techniques.

Assessing balance

To select between different matching algorithms and to assess covariate balance after matching, this study compared matching models along a variety of dimensions. First, individual covariate balance was assessed across samples by looking at the standardised difference in means across treatment and control groups both before and after matching. This standardised difference is the difference in group averages over the square root of the average of the sample variances. If samples are balanced, this difference should be small and matching should reduce this standardised difference as compared to the unmatched samples.

In addition, this study performed t-tests to assess whether differences across treatment and control groups were statistically significant. If balance is achieved with PSM, differences between treatment and control groups should be negligible and therefore should not be significantly different from zero.

In this context, the variance ratios of covariates of treated over control measures was also assessed. If there is perfect balance across samples, then covariates should be distributed equally and hence this ratio should be equal to one.

All of these measures give an indication of whether specific individual covariates are balanced across treatment and control groups. To assess overall variance, this study used two statistics that summarise covariate balance in the sample at hand: Rubin's B and Rubin's R. Rubin's B reflects the absolute standardised difference of the means of the propensity score in the treated and control groups (unmatched and matched). Rubin's R is the ratio of the treated to control variances of the propensity scores. Rubin (2001) suggests that the value of B should lie below 25 and that R should lie between .5 and 2 for overall balance to be sufficient. Together, Rubin's B and Rubin's R provide an informative indication of the trade-off between bias and variance across the treatment and control groups, as it changes before and after the matching procedure. However, individual-level balance should always be assessed as the overall balance is only an approximation of goodness of fit.

Matching procedures were implemented using the psmatch2 package in Stata (14.1) and balancing tests were carried out using the pstest package, which provides the results for all of the statistics mentioned above.[13]

Finally, the distribution of propensity scores was also analysed graphically. Ideally, propensity scores should be distributed equally across treatment and control groups. Very skewed/diverging distributions could be an indication that balance has not been achieved successfully. The visual distribution of propensity scores was therefore taken into account in selecting the preferred estimation model for the impact analysis.

Results of balancing assessments are presented in Section 6.7 of this report.

What treatment effect does PSM estimate?

It is important to emphasise that the PSM approach used in this study works by looking for control units that can be compared to treatment units, and not the other way round. This means that it is assumed that treatment units are a given and control units need to be identified. Through finding matches for the treatment units (pupils and teachers in EQUIP-T schools) in the pool of control units (pupils and teachers in non-EQUIP-T schools), the resulting estimate of the treatment effect

---

[13] See http://fmwww.bc.edu/repec/bocode/p/pstest.html for details.

is therefore the Average Treatment Effect on the Treated (ATT). Extrapolating this estimate beyond the population for which the treatment sample is representative is not immediately possible.

Considerations for applying PSM in the context of this evaluation

The following sections further outline the approach to employing PSM in the context of the present evaluation. In particular, it is important to emphasise that PSM was used as the core strategy to answer questions of programme impact at baseline and endline. However, initial analyses conducted within the context of the baseline report (OPM 2015b, Volume II, Section 3.5) indicated that perfect balance could not be achieved for all indicators. In particular, some outcome indicators showed significant difference between treatment and control groups, despite showing appropriate covariate balance, at baseline. This imbalance in outcomes was surprising, given that outcome measures were not supposed to have been influenced by the programme yet.

In order to address this issue, the current study combined PSM with a Difference-in-Difference approach. Section 6.3 briefly explains the general theoretical background to this method, while Section 6.4 explains in more detail how the two approaches were combined in the present context.

6.3   Difference-in-Differences

Difference-in-Differences (DID) is an approach that exploits the fact that data from the same treatment and control schools was collected at two points in time, at baseline and at midline. The idea behind this approach is quite straightforward: it compares data from treatment and control schools both at baseline and midline. This happens separately first. Then, in a second step, these baseline and midline comparisons are compared to each other. If, for example, the difference at baseline between treatment and control was smaller than at midline, this would indicate that the treatment has had an effect on treatment observations.

Figure 3 below exemplifies this logic.

In the present case, the comparisons at baseline and midline in the first step are not simple comparisons of descriptive statistics, but PSM estimations of any statistical significant differences between treatment and control groups. Estimates from these are then, in a second step, compared to each other across time. The key impact estimates presented in this report are the results of this double difference of PSM estimates.

Figure 3:   Visual representation of Difference-in-Differences analysis

Notes: Hypothetical data.
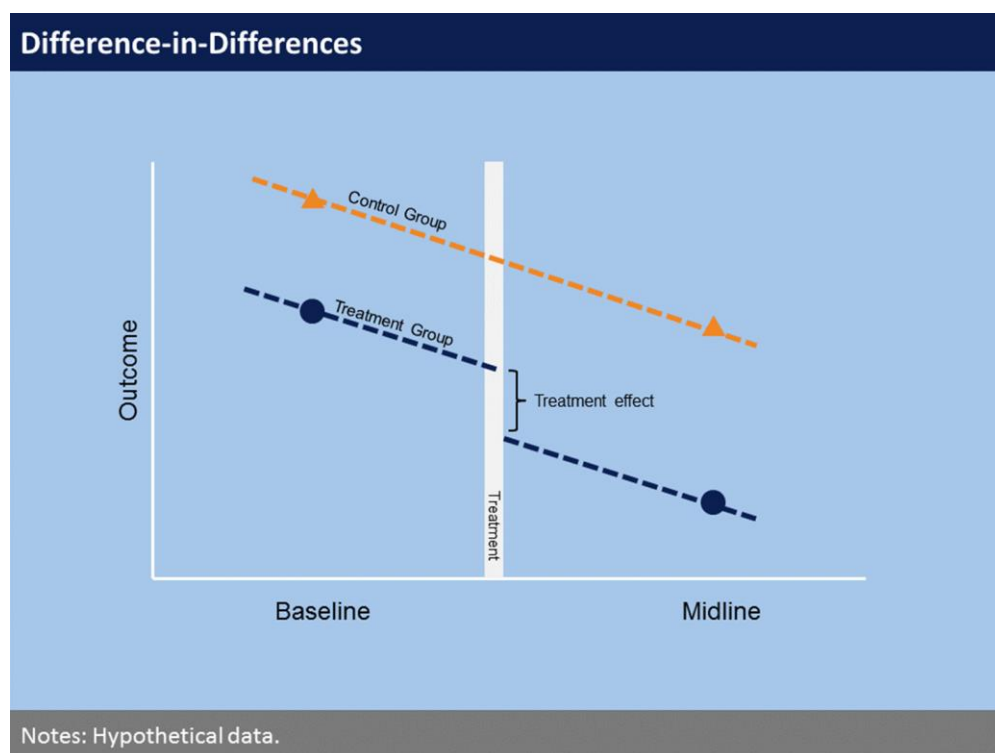
The key assumption that needs to hold for DID to identify programme effects is that, as can be seen in Figure 3 above, without the treatment (i.e. the EQUIP-T intervention) the difference between control and treatment groups at the second time-period (i.e. midline) would have been the same as in the first time-period (i.e. baseline). This is referred to as the parallel trend assumption.

In the present case, this means that without the treatment, imbalances remaining after PSM would be the same at baseline and at midline. Note that this means that such imbalances must be assumed to be time-invariant, i.e. constant across time. Taking the second difference across time removes such baseline imbalances from the estimation, which hence allows to isolate and robustly infer programme impact.

Importantly, for panelled observations, this also includes time-invariant unobservable characteristics that might be correlated to the outcome measure and the treatment status for panelled observations. In the present case, this means that any such school-level characteristics are also controlled for. This increases the robustness of findings, because PSM alone cannot control for unobservable characteristics driving selection bias. (See Section 6.2)

Therefore, combining DID with PSM helps to control for remaining imbalances that may exist between treatment and control groups after matching. Assuming that these are time-invariant, taking the difference between matched comparisons at baseline and at midline, allows to isolate with confidence the programme impact on beneficiaries (i.e. teachers and pupils in this case).

6.4   Combining DID and PSM

In this study, two different approaches have been used to combine PSM with DID:

1. Directly comparing ATT estimates at ML and BL across time.
2. Matching treatment observations across time to construct a pseudo panel of treatment observations and to construct an overall ATT estimate using this pseudo panel only.

The first approach was to take a direct difference of baseline and midline estimations of ATTs derived from PSM at baseline and midline. Essentially, this amounts to comparing two estimated treatment coefficients with each other. In theory, ATT estimates at baseline should be close to zero – because EQUIP-T had not started at that time yet. However, as described above and as can be seen in Section 6.7, this was not always the case, despite good balancing performance of models at baseline. Taking into account the ATT estimate at midline therefore means that the overall impact of EQUIP-T is defined as the difference that EQUIP-T made in the estimated ATT at midline, compared to the baseline estimate:

$$ATT_{overall} = ATT_{midline} - ATT_{baseline} \cdot (2)$$

Of course, the main goal is to conduct inference on this estimate, i.e. to see whether the overall ATT estimate is different from zero or not. Test statistics for the estimate defined in (2) are calculated using the formula for comparing coefficient estimates presented in Paternoster et al. (1998). Using this test statistic, this study then calculates whether the estimated ATT is significantly different from zero or not from a statistical point of view. Note that all standard errors for the midline and baseline ATT used are based on bootstrapping procedures for PSM estimates. (See section 6.6 on why standard errors for PSM are bootstrapped.)

The **second approach** is a robustness check where additional matching is used to create a 'pseudo panel' of treatment observations (i.e. teachers and pupils in EQUIP-T schools) across time, given that these have not been panelled and were surveyed as repeated cross-sections. Figure 3 depicts this process graphically.

In a first step, treatment observations from teacher and pupil samples are uniquely matched across the two time periods. This is done using a Nearest Neighbour PSM approach without replacement. This means that for each treatment observation at baseline a unique comparator is found at midline. See Section 6.2 on how such a matching approach can be implemented.

For this 'pseudo panel' of treatment observations, values obtained for their respective matched comparisons at baseline and midline are then used to calculate differences between estimated control group and treatment group individuals at baseline and at midline separately, using the same PSM models as in the main estimations. Note that kernel matching at baseline and midline provides, for each treatment observation, an appropriate estimated counterfactual value based on the PSM estimation. This value is used to calculate the first difference between treatment observations and counterfactuals, as part of the double differencing approach underpinning the Diff-in-diff analysis. In a final step, those differences are then compared across baseline and midline for the 'pseudo panel'. The average of this double difference for the pseudo panel is the estimated overall ATT. By implementing this approach, this study follows a suggestion by Blundell and Costa Dias (2000, p. 451). This study is likely to represent the first practical application of this PSM with Diff-in-diff procedure for a repeated cross-section, in an education evaluation of teachers and pupils.

The key difference to the first approach is that this double differencing is implemented only across treatment observations that are similar to each other, as they have been matched one-to-one in the first step.

Figure 4:   Visual representation of second PSM with DID combination

Building a pseudo panel of treatment observations for the PSM with DID estimation
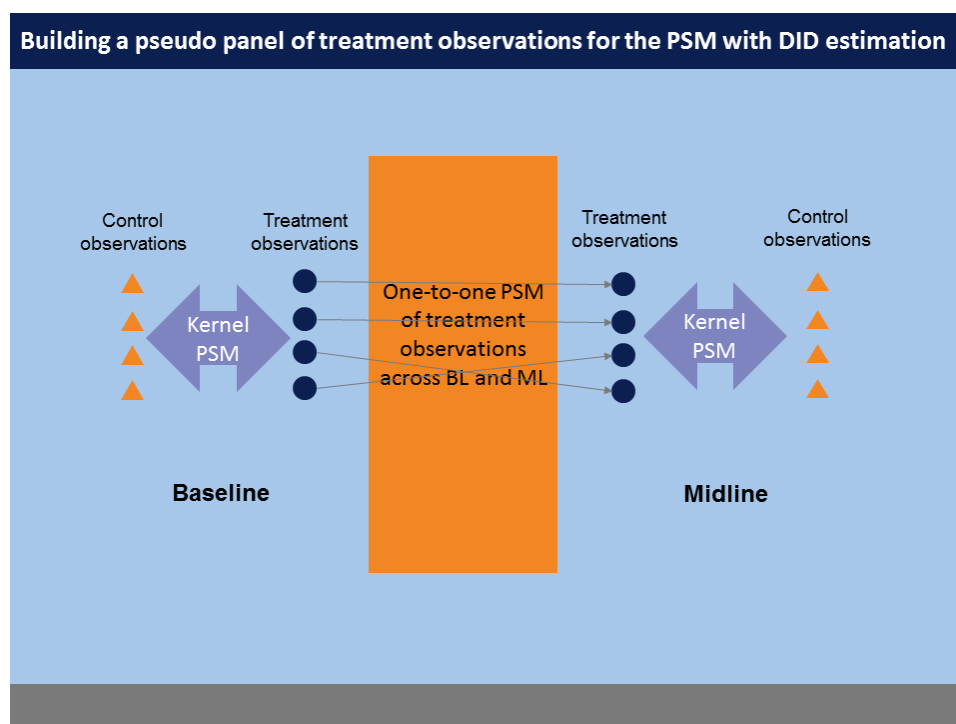
6.5   How results are presented in Volume I

In Volume I, headline results are presented in a visual form, with an explanation underneath each graph. These headline results are the results of the first DID PSM combination approach. See Figure 4 below for an example. Each graph shows point estimates for treatment effects (ATT) on outcome indicators and 95% confidence intervals for these effects. This means that the probability for the true treatment estimate to fall within this area is 95%.

Outcome indicators used in this evaluations are mainly proportions. This means that estimates of treatment effects are given in percentage point changes of these proportions. For example, if the ATT estimate on the proportion of pupils in the bottom performance band of Kiswahili in treatment schools is -0.03, this means that EQUIP-T has reduced this proportion by three percentage points, compared to a counterfactual of no EQUIP-T package and some alternative teacher INSET training. Equivalently, this can be expressed as a decrease of three percentage points in the probability of pupils from treatment schools to fall in this bottom performance band. When confidence intervals of such estimates do not overlap with zero, then this is an indication that this treatment effect is truly different from zero. This zero value is indicated using a red line in the graphs.

Figure 5:   Impact of EQUIP-T on pupil learning
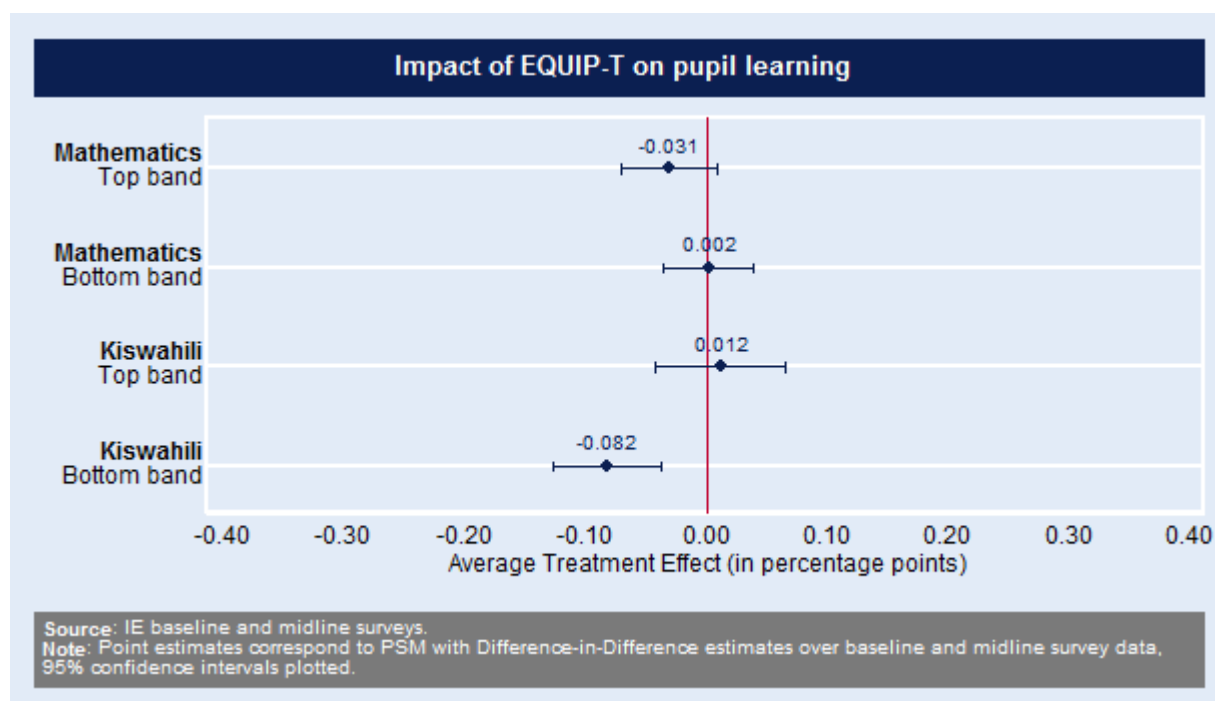
As explained above, the ATT estimates shown here are the results of the first approach to implement PSM DID estimations, and thus take into account information both from the baseline and the midline data. Comparing Figure 4 to Figure 5 can help to understand this. Figure 5 shows estimates of averages of the treatment group and of matched counterfactuals at baseline and at midline. Note that the control estimates are not simple descriptive statistics – they are the averages of counterfactual observations constructed using PSM. The PSM DID estimates presented in Figure 4 correspond to the double difference of the averages presented below. The first difference at midline is 0.214 - 0.162 = 0.052. The same difference at baseline is 0.389 - 0.254 = 0.135. The double difference estimate is 0.05-0.14 = -0.083. This corresponds to the ATT estimate presented in Figure 4. When looking at the graph below, one can see that the difference between EQUIP-T and control schools has effectively decreased over time – this decrease in difference is the ATT estimate and is due to EQUIP-T.

Figure 6:   Example PSM comparisons

Source: IE baseline and midline surveys.
Note: Point estimates correspond to mean values of matched outcomes at baseline and midline, 95% confidence intervals plotted.
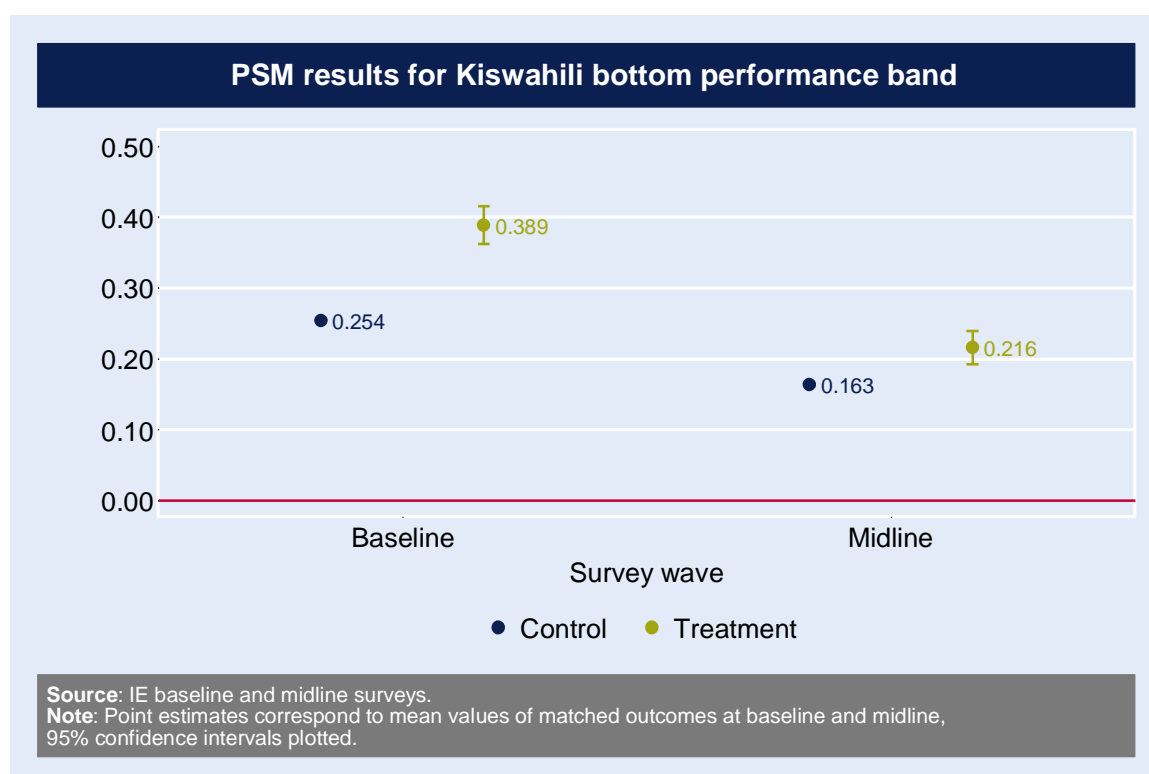
6.6   Caveats - Addressing weaknesses in the analysis

Four key caveats related to the present estimation strategy need to be mentioned here. First, PSM only controls for observable characteristics that cause selection bias. This is a problem for any impact identification strategy that relies on controlling only for factors (variables) that can be observed in the data – not only PSM. PSM helps addressing this by allowing for extensive balancing checks after matching, which can provide substantial evidence for the fact that balance is achieved across a large variety of characteristics and – by implication – is likely to also extent to unobservables. In this study, such extensive balancing checks were implemented. Results are presented in Section 6.7 below. In addition, as explained above, the DID strategy implemented in the present case helps to control for remaining imbalances that may be due to time-invariant unobservable variables.

Second, DID helps to deal with time-invariant imbalances, but not time variant ones. This means that only time-invariant imbalances that remain after PSM would be controlled for, in contrast to imbalances that vary over time. In the present case, this is addressed by extensive balancing tests, which show little remaining covariate imbalance in general after PSM, by showing that results are robust to a variety of different PSM specifications, and by showing that results are robust to two separate DID strategies used. Together, this evidence implies that results are robust, remaining imbalances are small, and results are unlikely to be sensitive to or to be driven by such imbalances – even if they were time variant.

Third, as discussed in Volume I of this ML report Chapter 1, INSET teacher training and SLM training for HTs have been implemented not only in EQUIP-T schools, but also in control schools. This means that the impact identified by the analysis is the effect that EQUIP-T as a package (including related trainings) has had on the outcome indicators compared to a counterfactual situation where in the same schools the alternative training from control schools would have been

implemented. In other words, the analysis measures the compounded impact of all EQUIP-T related interventions over the potential effect of other training initiatives that took place in both treatment and control schools. This still allows us to identify the marginal impact attributed to EQUIP-T and thus its added value. Attributing impact to EQUIP-T would become more problematic if the LANES INSET training is repeated and expanded between the midline and endline stages of the evaluation. In this case, the marginal impact of EQUIP-T would be more difficult to capture with confidence. This is explained in more detail in section **Error! Reference source not found.** on risks to the evaluation.

Finally, calculating standard errors of estimated treatment effects using PSM methods is not straightforward. As Caliendo and Kopeinig (2005, p. 18) put it, 'The problem is that the estimated variance of the treatment effect should also include the variance due to the estimation of the propensity score, the imputation of the common support, and possibly also the order in which treated individuals are matched'. These estimations increase the variation of the treatment effect estimates over and above normal sampling variation. In the literature, there is no consensus on how to take this into account.

A popular approach to solve this problem is to bootstrap standard errors for the estimated treatment effect (see Lechner 2002). Each bootstrap draw re-estimates both the first and second stages of the estimation. This produces N bootstrap samples for which the ATT is estimated. The distribution of these means approximates the true sampling distribution, and therefore the standard errors of the population mean (Caliendo and Kopeinig 2005, p.18). This study followed this approach and implemented bootstrapping, using 200 repetitions, to estimate the standard errors of the estimated treatment effects. Note that, for the sake of completeness, this report shows both the bootstrapped and the non-bootstrapped standard errors below.

It is also important to note that there is no clear direction in which estimated standard errors should change due to bootstrapping. On the one hand, the additional variation taken into account should increase standard errors. On the other, bootstrapping generally makes estimates more precise, which tends to decrease standard errors. Overall, the direction of the change is not uniform. In fact, the results show that, with bootstrapping, standard errors in some instances are smaller and in some larger than without bootstrapping.

6.7   Results

This section presents the results obtained from applying PSM to EQUIP-T baseline and midline data. In the following paragraphs, the balancing results, the ATT estimates and the PSM-DID estimates described for all impact indicators for the main strategy and the robustness check[14] . The following indicators were analysed in the context of this evaluation:

Table 7:   Impact indicators for PSM-DID estimation

Impact area	Impact indicators	Sample for the impact evaluation

---

[14] It is important to highlight the fact that a large range of results where produced in the course of the analysis across a range of different models, including varying levels of trimming and bandwidth size for the kernel matching algorithm. This extensive investigation of alternative specifications provided the opportunity to select the most appropriate and robust estimation strategies for which results are presented in this report. At the same time, consistency or inconsistency in the direction and significance of results emerging from this range of models help determine whether any findings on impact (or lack of thereof) can be considered conclusive or yet inconclusive.

Pupil learning	Proportion of pupils in the bottom performance band of the interval scale for Mathematics	Standard 3 pupils who were assessed
	Proportion of pupils in the top performance band of the interval scale for Mathematics	
	Proportion of pupils in the bottom performance band of the interval scale for Kiswahili	
	Proportion of pupils in the top performance band of the interval scale for Kiswahili	
Teacher absenteeism	Proportion of teachers who were absent on the day of the survey	All teachers (from roster)
	Proportion of teachers present on the day of the survey, timetabled to teach before lunch and absent from the classroom	
School leadership and management	Proportion of teachers who report participation in performance appraisals	Interviewed teachers of Standards 1-3

For each of the outcome variables, this study implemented two PSM DID strategies, one main strategy and a robustness check outlined in Section 6.4.

Presentation of results

For each outcome variable, three sets of results are presented in this volume: (a) the second stage results, (b) the propensity score matched outcomes at baseline and midline, and (c) the PSM-DID estimates. The following paragraphs use the example of Figure 7 to explain the interpretation of results in detail.

First, the second stage results for the main strategy are presented, as illustrated in Figure 7 for the indicator on top performance band for Mathematics. The figure is divided into two panels; the top panel and the bottom panel show baseline and midline results respectively. The format for each panel is as follows:

- The first graph on the left-hand side indicates how individual variables balance before and after matching. The x-axis displays the standardised bias, which is the percentage difference of the sample means in the treated and non-treated (unmatched or matched) subsamples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Rosenbaum and Rubin 1985). In Figure 7 below, for example, the unmatched samples display large imbalances with standardised bias being present across many of the covariates of interest. However, once matching takes place, the standardised imbalances are diminished.

- The second graph, on the right-hand side, shows the distribution of propensity scores across treatment and control groups. This graph visually confirms that, after dropping observations that are off common support, both treatment and control groups contain observations with propensity scores across the full range of the distribution, which is an indication for overall balance. Although the distributions of propensity scores across treatment and control groups would ideally be symmetric, the presence of some level of skewness does not put at risk the estimation procedure, as indicated by the balance achieved for each covariate and the overall values of Rubin's R and B after matching.

- The remaining rows on the right hand side display information related to the PSM model. The bandwidth and level of trimming for the optimal PSM model can be found in the first two rows. For example, the optimal model has a bandwidth of 6 and a trimming value of 10 for the baseline sample in Figure 6. This is then followed by the number of observations on common support in the next row, and then the Rubin's R and Rubin's B values both before and after matching. Generally, a Rubin's B score under 25 after matching is desirable, whilst a Rubin's R score between 1 and 1.25 is the preferred range after matching (Rubin 2001). The unmatched samples are particularly unbalanced; for instance, the Rubin's B for the baseline sample and the midline sample is 85.47 and 70.87 respectively. However, the Rubin's B scores after matching, which are all below 25, show how matching removes the previous imbalances.

- Finally, the remaining rows on the right hand side indicate the ATT for each corresponding survey wave and the associated standard errors. Given that it is not definitively clear how to produce standard errors for PSM, both bootstrapped and non-bootstrapped standard errors are presented for robustness purposes. (See Section 6.6 for more detail on this.)

Proportion of pupils in the top performance band for Mathematics

Figure 7:   Mathematics top band: Second stage results (Main strategy)

Standardized bias across covariates and ATT	Balancing	
Baseline		

ATT	0.00
SE (bootstrapping)	(0.011)
SE (no bootstrapping)	(.01)

Midline

Bandwidth		2
Trimming		13
N on common support		2576
Rubin's B	[before	85.47
Rubin's R	matching]	0.48
Rubin's B	**[after**	**25.51**
Rubin's R	**matching]**	**1.09**

Bandwidth	4
Trimming	13

N on common support		2505
Rubin's B	[before matching]	70.87
Rubin's R		1.09

ATT	-0.03	Rubin's B	[after matching]	22.31
SE (bootstrapping)	(0.018)	Rubin's R		0.94
SE (no bootstrapping)	(0.014)			

Second, the mean values of the matched outcome and associated confidence intervals at baseline and midline for the treatment group and the control group are plotted. An example can be seen in Figure 7 for top performance band of Mathematics. For the treatment group, the mean of the outcome variable is plotted for observations on common support. For the control group, the mean of the counterfactual outcome estimated by the matching algorithm is plotted here.

Figure 8:   Mathematics top band: Matched outcome at baseline and midline
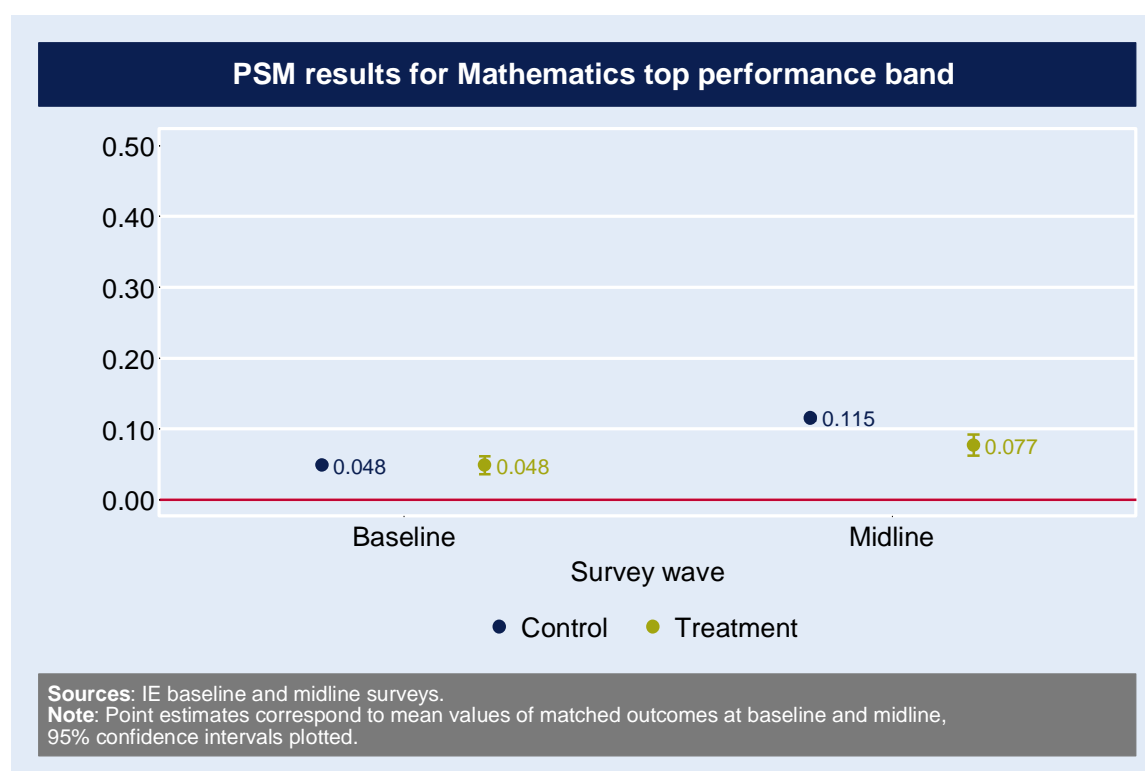
Finally, the PSM Differences-in-Differences estimate for both the main strategy and the robustness check are presented, along with the associated bootstrapped and non-bootstrapped p-values. See Table 2 as an example of how the overall impact result should be interpreted across

the two strategies. In that table, the PSM-DD estimate from the robustness check shows a statistically significant negative trend in EQUIP-T schools, although this finding is not confirmed by the main strategy which fails to detect a similarly significant negative trend.

Whilst the PSM-DD estimate from the robustness check shows a statistically significant negative trend in EQUIP-T schools, this finding is not confirmed by the main strategy where the analysis fails to detect a similar trend.

Table 8:   Mathematics top band: PSM-DID estimate

	Main strategy	**Robustness check**
PSM-DID estimate	**-0.03**	**-0.04**
P-value (bootstrapping)	**(0.13)**	**(0.003)**
P-value (no bootstrapping)	**(0.07)**	**(0.003)**

The balancing results for the robustness check strategy - where treatment observations across the two survey waves are matched - are also summarized at the end for each outcome indicator, as illustrated in Table 3. This table shows that the balancing properties for this matching process concerning this particular indicator were not ideal – note that Rubin's R is above 25. Although this strategy does not confirm the finding from our main strategy, this cannot lead us to change our overall conclusion that EQUIP-T did not have a significant impact on this outcome indicator. This is due to the weakness in the robustness strategy identification in this specific case.

Table 9:   Mathematics top band: Balancing results (Robustness check strategy)

Balancing results from matching treatment observations across baseline and midline		
Caliper		.4
N for common support		1586
Rubin's B Rubin's R	[before matching]	89.31 1.27
Rubin's B Rubin's R	[after matching]	28.4 0.9

Proportion of pupils in the bottom performance band for Mathematics

Figure 9:   Mathematics bottom band: Second stage results (Main strategy)

Standardized bias across covariates and ATT	**Balancing**
Baseline	

Bandwidth	4
Trimming	10
N on common support	2519

| Rubin's B | [before | 102.72 |
| Rubin's R | matching] | 0.88 |
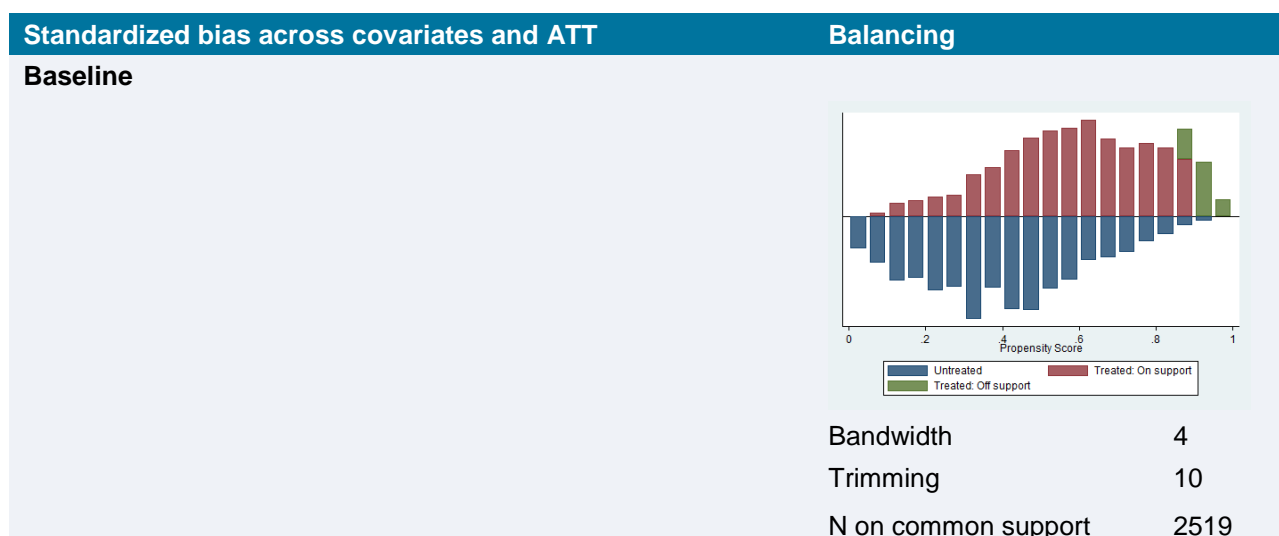
ATT	0.04
SE (bootstrapping)	(0.014)
SE (no bootstrapping)	(0.015)

Midline

| **Rubin's B** | **[after** | **20.6** |
| **Rubin's R** | **matching]** | **1.36** |
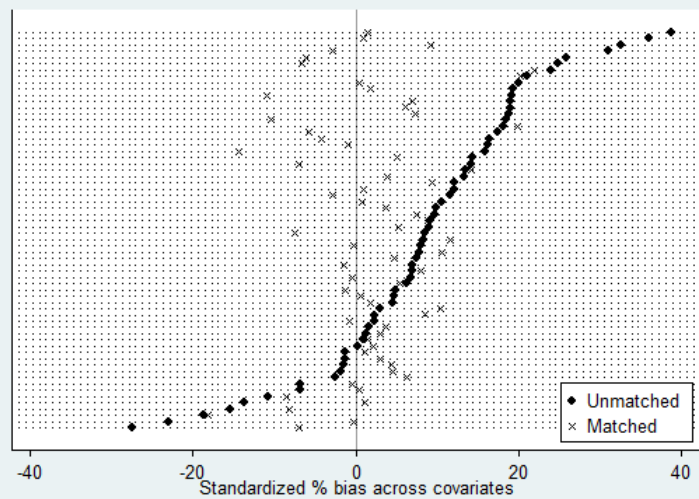
ATT	0.05
SE (bootstrapping)	(0.012)
SE (no bootstrapping)	(0.017)
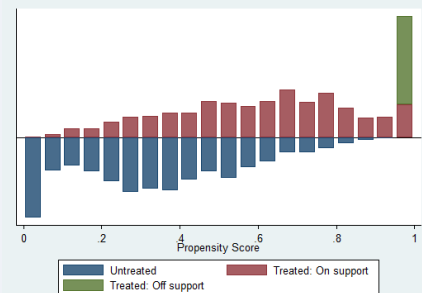
Bandwidth	6
Trimming	15
N on common support	2605
Rubin's B	[before	50.77
Rubin's R	matching]	7.82
Rubin's B	[after	25.83
Rubin's R	matching]	0.3

Figure 10:   Mathematics bottom band: Matched outcome at baseline and midline

PSM results for Mathematics bottom performance band

Sources: IE baseline and midline surveys.
Note: Point estimates correspond to mean values of matched outcomes at baseline and midline, 95% confidence intervals plotted.
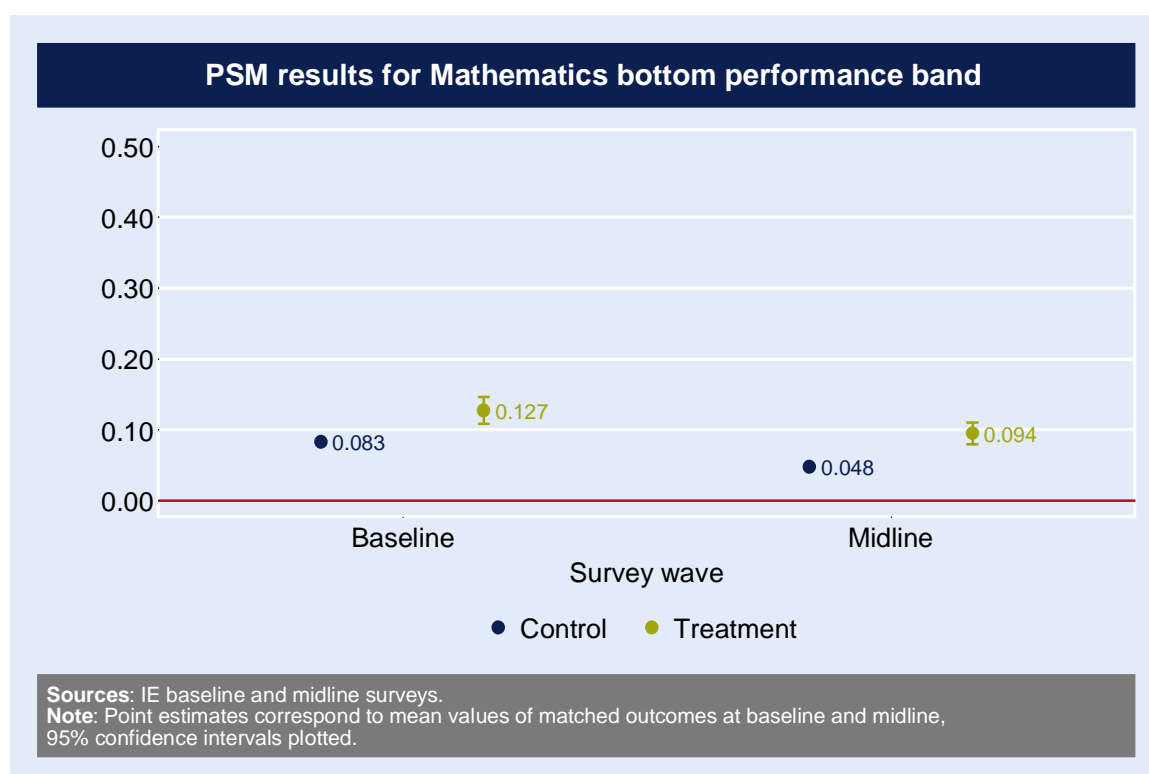
Figure 10 above shows that the PSM estimates point to an overall decrease in the proportion of pupils in the bottom performance band for Mathematics, but without much difference in this trend across treatment and comparison schools. As can be seen in Table 10 below, this means that the study does not find any evidence of a statistically significant impact of EQUIP-T on the proportion of pupils in the bottom performance band for Mathematics. The main strategy and the robustness check are consistent with each other with regards to this assessment.

Table 10:   Mathematics bottom band: PSM-DID estimate

	Main strategy	**Robustness check**
PSM-DID estimate	0.002	-0.001
P-value (bootstrapping)	(0.92)	(0.50)
P-value (no bootstrapping)	(0.93)	(0.50)

Table 11 below presents results on the balancing properties of the robustness check strategy. As can be seen in the 'after matching' row, balancing is not ideal for treatment observations across time.

Table 11:   Mathematics bottom band: Balancing results (Robustness check strategy)

Balancing results from matching treatment observations across baseline and midline		
Caliper		0.4
N for common support		1844
Rubin's B **Rubin's R**	[before matching]	85.89 1.04
Rubin's B Rubin's R	**[after matching]**	**28.09** **0.81**

Proportion of pupils in the top performance band for Kiswahili

Figure 11: Kiswahili top band: Second stage results (Main strategy)

Standardized bias across covariates and ATT	Balancing

Baseline

ATT	**-0.04**
SE (bootstrapping)	**(0.018)**
SE (no bootstrapping)	**(0.016)**

Bandwidth		6
Trimming		10
N on common support		2564
Rubin's B	[before	87.75
Rubin's R	matching]	0.5
Rubin's B	**[after**	**23.96**
Rubin's R	**matching]**	**0.97**

Midline

ATT	-0.03
SE (bootstrapping)	(0.023)
SE (no bootstrapping)	(0.021)

Bandwidth		6
Trimming		10
N on common support		2643
Rubin's B	[before	71.97
Rubin's R	matching]	1.1
Rubin's B	[after	28.83
Rubin's R	matching]	1.47

Figure 12: **Kiswahili top band: Matched outcomes at baseline and midline**

Sources: IE baseline and midline surveys.
Note: Point estimates correspond to mean values of matched outcomes at baseline and midline, 95% confidence intervals plotted.
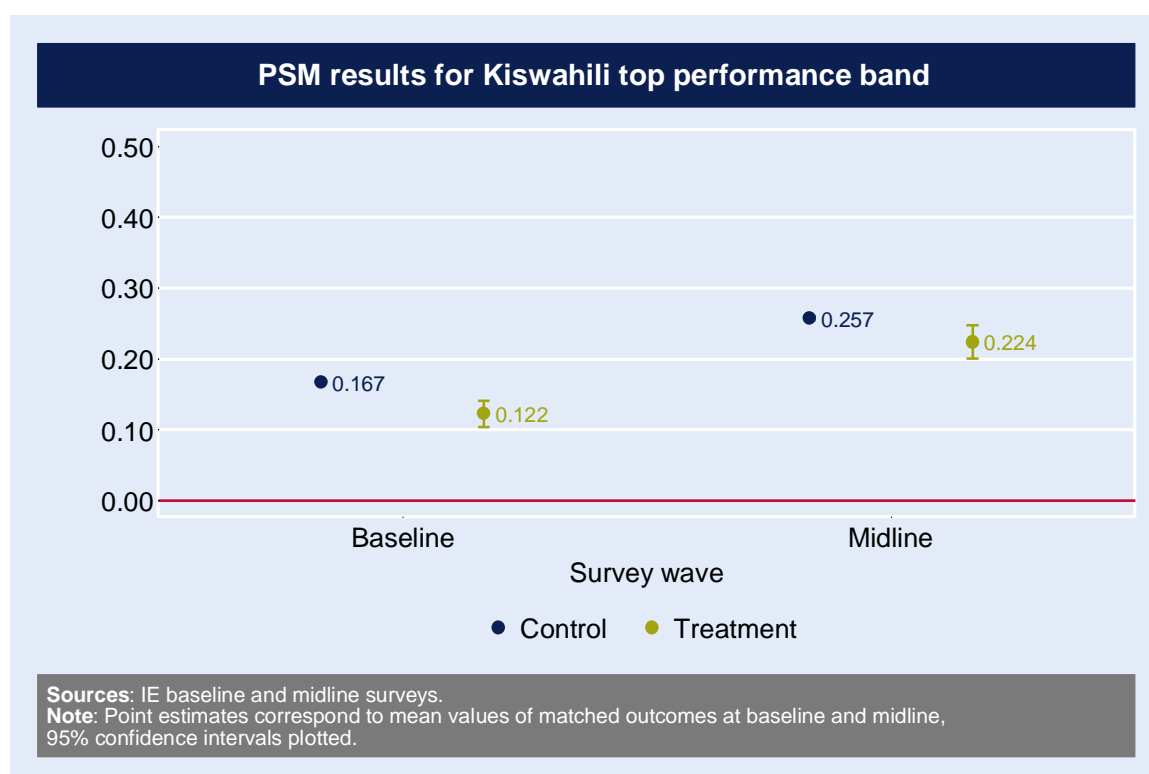
Whilst both strategies show a positive change in the proportion of pupils in the top performance band for Kiswahili, this result is not statistically significant and therefore, the analysis is unable to provide a positive assessment on the impact of EQUIP-T on this indicator.

Table 12: **Kiswahili top band: PSM-DID estimate**

	Main strategy	**Robustness check**
PSM-DID estimate	0.012	0.02
P-value (bootstrapping)	(0.67)	(0.28)
P-value (no bootstrapping)	(0.68)	(0.28)

Table 13: **Kiswahili top band: Balancing results (Robustness check strategy)**

Balancing results from matching treatment observations across baseline and midline		
Caliper		0.3
N for common support		1630
Rubin's B Rubin's R	[before matching]	88.5 1.18
Rubin's B **Rubin's R**	**[after matching]**	**24.63** **0.43**

Proportion of pupils in the bottom performance band for Kiswahili

Figure 13: Kiswahili bottom band: Second stage results (Main strategy)

Standardized bias across covariates and ATT	Balancing

Baseline

ATT	0.13
SE (bootstrapping)	(0.02)
SE (no bootstrapping)	(0.02)

Bandwidth		6
Trimming		8
N on common support		2641
Rubin's B	[before	99.25
Rubin's R	matching]	0.47
Rubin's B	[after	18.98
Rubin's R	matching]	1.5

Midline

ATT	0.05
SE (bootstrapping)	(0.016)
SE (no bootstrapping)	(0.018)

Bandwidth		6
Trimming		10
N on common support		2577
Rubin's B	[before	68.57
Rubin's R	matching]	0.81
Rubin's B	[after	26.83
Rubin's R	matching]	1.33

Figure 14:   Kiswahili bottom band: Matched outcomes at baseline and midline

PSM results for Kiswahili bottom performance band

Source: IE baseline and midline surveys.
Note: Point estimates correspond to mean values of matched outcomes at baseline and midline, 95% confidence intervals plotted.
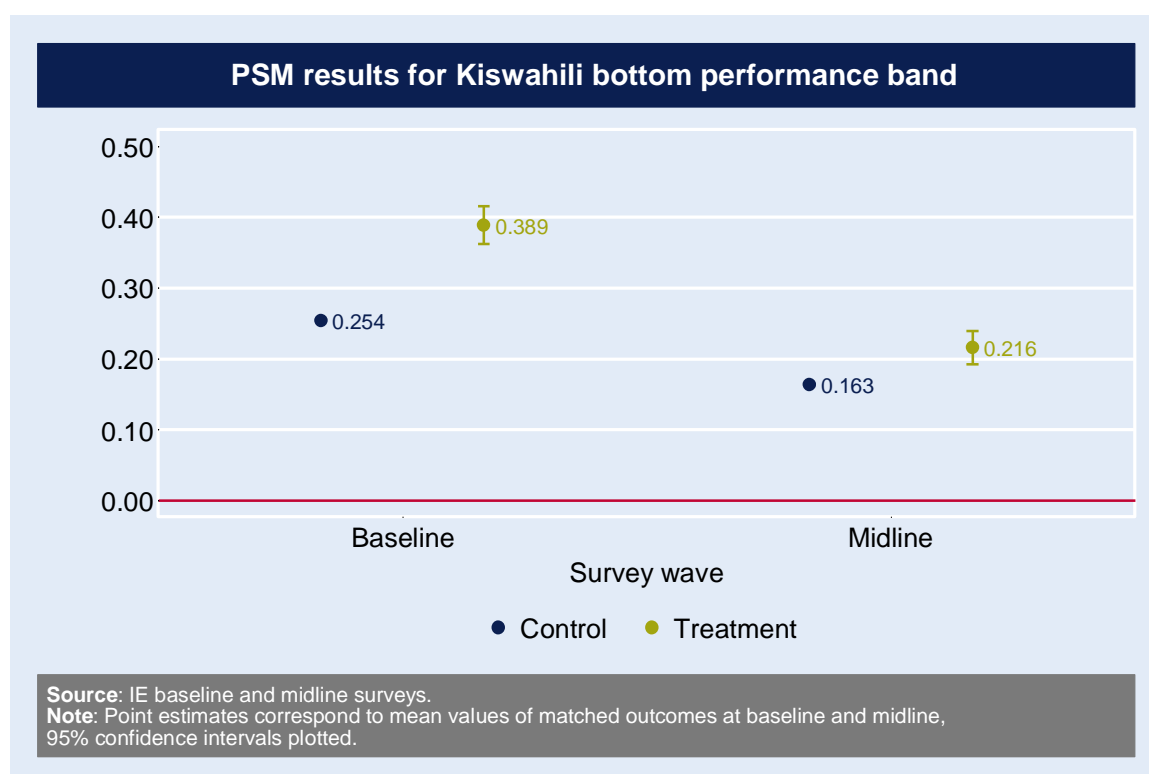
Figure 14 above shows that the PSM analyses at BL and ML point to a decreasing gap between treatment and comparison schools in terms of pupils who are in the bottom performance band of Kiswahili. This means that the overall PSM DID analysis finds strong evidence that EQUIP-T has reduced the proportion of pupils in the bottom performance band for Kiswahili in programme schools. See Table 14 below for this. These results remain strong and highly significant across both the main strategy and the robustness check.

Table 14:   Kiswahili bottom band: PSM-DID estimate

	Main strategy	Robustness check
PSM-DID estimate	-0.08	-0.07
P-value (bootstrapping)	(0.00)	(.00)
P-value (no bootstrapping)	(0.00)	(0.001)

The balancing results for the robustness check matching across time for treatment observations, presented below, show that for this outcome indicator balancing after matching is within acceptable ranges. This further strengthens the findings presented above, that EQUIP-T has significantly reduced the proportion of children in the bottom performance band for Kiswahili in treatment schools, compared to a counterfactual situation without EQUIP-T.

Table 15:   Kiswahili bottom band: Balancing results (Robustness check strategy)

Balancing results from matching treatment observations across baseline and midline		
Caliper		0.4
N for common support		1798
Rubin's B **Rubin's R**	[before matching]	91.49 1.18
Rubin's B Rubin's R	**[after matching]**	**24.99** **1.01**

Teacher school absenteeism

Figure 15:   Teacher school absenteeism: Second stage results (Main strategy)

Standardized bias across covariates and ATT	Balancing

Baseline

ATT	0.03	
SE (bootstrapping)	(0.023)	
SE (no bootstrapping)	(0.018)	

Bandwidth		2
Trimming		8
N on common support		1670
Rubin's B Rubin's R	[before matching]	35.62 0.04
Rubin's B **Rubin's R**	**[after matching]**	**19.11** **1.18**

Midline

Bandwidth		4
Trimming		10
N on common support		1836
Rubin's B Rubin's R	[before matching]	62.21 2.03
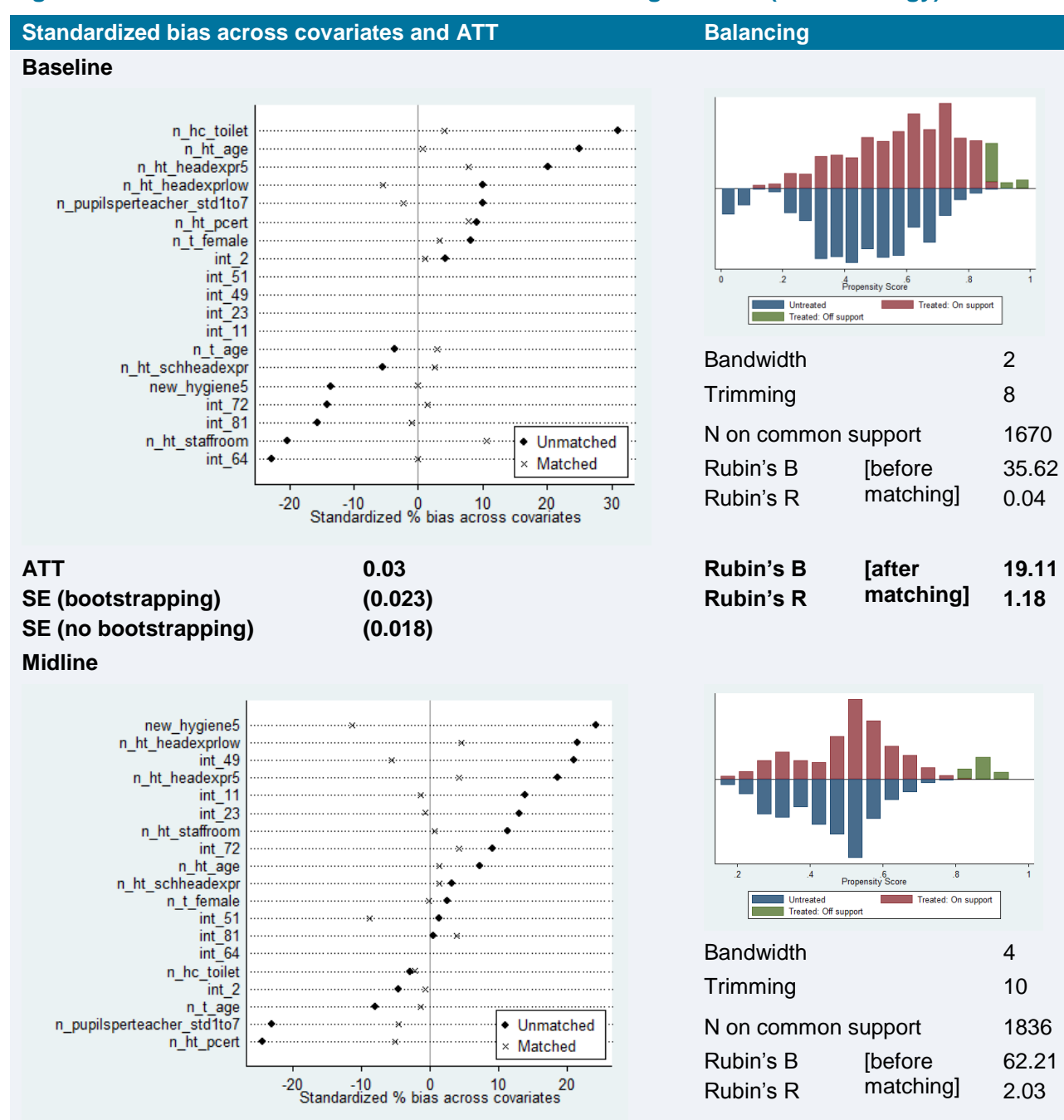
ATT	0.02	Rubin's B	[after matching]	18.84
SE (bootstrapping)	(0.015)	Rubin's R		1.05
SE (no bootstrapping)	(0.015)			

Figure 16:   Teacher school absenteeism: Matched outcomes at baseline and midline

PSM results for teacher school absenteeism

Baseline: Control 0.106, Treatment 0.141
Midline: Control 0.106, Treatment 0.122

Survey wave

● Control   ● Treatment

Sources: IE baseline and midline surveys.
Note: Point estimates correspond to mean values of matched outcomes at baseline and midline, 95% confidence intervals plotted.
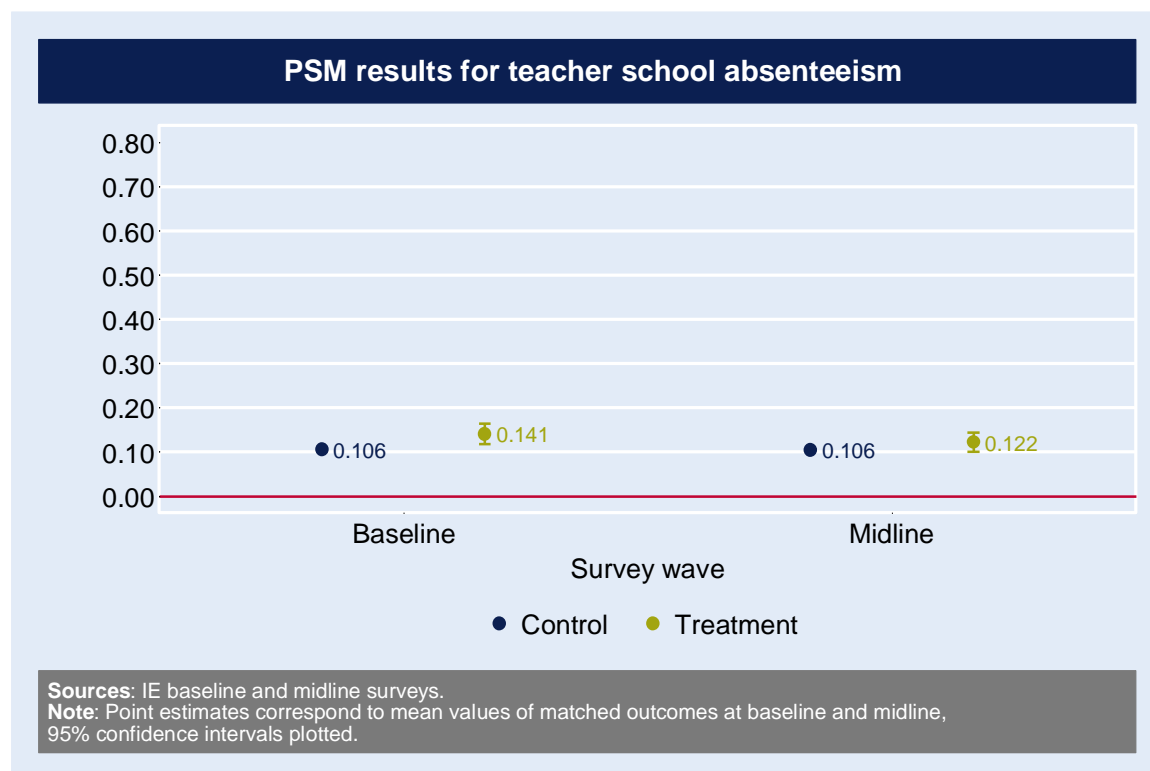
The analysis does not find strong evidence for an impact in terms of the proportion of teachers who are absent from school on the day of the survey. As shown in Table 16, both models display positive trends on this indicator, but the results are not significant in the main strategy and only weakly significant in the robustness check. Therefore, these results do not point conclusively towards an impact of EQUIP-T on this absenteeism indicator.

Table 16:   Teacher school absenteeism: PSM-DID estimate

	Main strategy	Robustness check
PSM-DID estimate	-0.018	-0.04
P-value (bootstrapping)	(0.50)	(0.022)
P-value (no bootstrapping)	(0.44)	(0.023)

Table 17:   Teacher school absenteeism: Balancing results (Robustness check)

Balancing results from matching treatment observations across baseline and midline		
Caliper		0.1
N for common support		1290
Rubin's B **Rubin's R**	[before matching]	80.14 1.12
Rubin's B Rubin's R	[after matching]	**24.22** **0.46**

Teacher classroom absenteeism

Figure 17: Teacher classroom absenteeism: Second stage results (Main strategy)

Standardized bias across covariates and ATT	Balancing

Baseline

ATT	0.08
SE (bootstrapping)	(0.035)
SE (no bootstrapping)	(0.032)

Bandwidth		6
Trimming		10
N on common support		1230
Rubin's B	[before	98.6
Rubin's R	matching]	1.09
Rubin's B	**[after**	**24.36**
Rubin's R	**matching]**	**1.43**

Midline

ATT	-0.04
SE (bootstrapping)	(0.033)
SE (no bootstrapping)	(0.031)

Bandwidth		2
Trimming		10
N on common support		1252
Rubin's B	[before	79.18
Rubin's R	matching]	1.18
Rubin's B	[after	20.84
Rubin's R	matching]	1.99

Figure 18:   Teacher classroom absenteeism: Matched outcome at baseline and midline

PSM results for teacher classroom absenteeism

Sources: IE baseline and midline surveys.
Note: Point estimates correspond to mean values of matched outcomes at baseline and midline, 95% confidence intervals plotted.
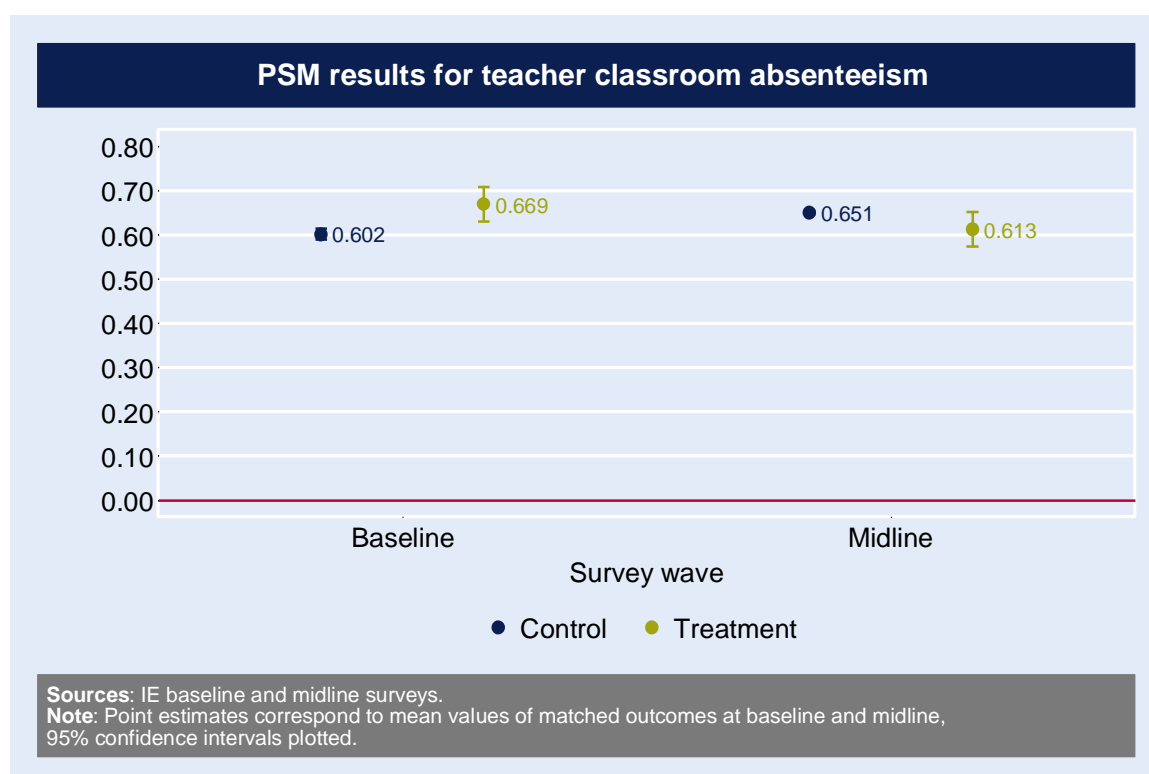
As shown in Figure 18 above, the PSM analyses indicate that at BL the proportion of teachers that were absent from classes was marginally higher in treatment schools than in comparison schools, whereas the opposite was true at ML. Consequently, and as can be seen below, the PSM DID analyses indicate a substantial reduction in the proportion of teachers absent from classes they are timetabled to teach before lunch as a result of EQUIP-T. These results stand up to scrutiny, performing consistently across both preferred and robustness check strategies.

Table 18:   Teacher classroom absenteeism: PSM-DID estimate

	Main strategy	**Robustness check**
PSM-DID estimate	**-0.115**	**-0.11**
P-value (bootstrapping)	**(0.02)**	**(0.00)**
P-value (no bootstrapping)	**(0.01)**	**(0.001)**

The table below shows that after matching the balancing properties for treatment observations across BL and ML are good, further strengthening the results presented above.

Table 19:   Teacher classroom absenteeism: Balancing results (Robustness check strategy)

Balancing results from matching treatment observations across baseline and midline		
Caliper		0.3
N for common support		938
Rubin's B **Rubin's R**	[before matching]	81.96 0.91
Rubin's B Rubin's R	**[after matching]**	**22.48** **0.66**

Proportion of teachers who report participation in performance appraisal

Figure 19: Teacher performance appraisal: Second stage results (Main strategy)

Standardized bias across covariates and ATT	Balancing

Baseline

ATT	0.03
SE (bootstrapping)	(0.042)
SE (no bootstrapping)	(0.041)

Bandwidth		4
Trimming		3
N on common support		634
Rubin's B	[before matching]	47.07
Rubin's R		0.08
Rubin's B	**[after matching]**	**16.98**
Rubin's R		**1.50**

Midline

ATT	0.11
SE (bootstrapping)	(0.032)
SE (no bootstrapping)	(0.033)

Bandwidth		4
Trimming		3
N on common support		707
Rubin's B	[before matching]	49.65
Rubin's R		12.83
Rubin's B	[after matching]	11.28
Rubin's R		0.79

Figure 20: **Teacher performance appraisal: Matched outcome at baseline and midline**

PSM results for teacher participation in performance appraisal

Sources: IE baseline and midline surveys.
Note: Point estimates correspond to mean values of matched outcomes at baseline and midline, 95% confidence intervals plotted.
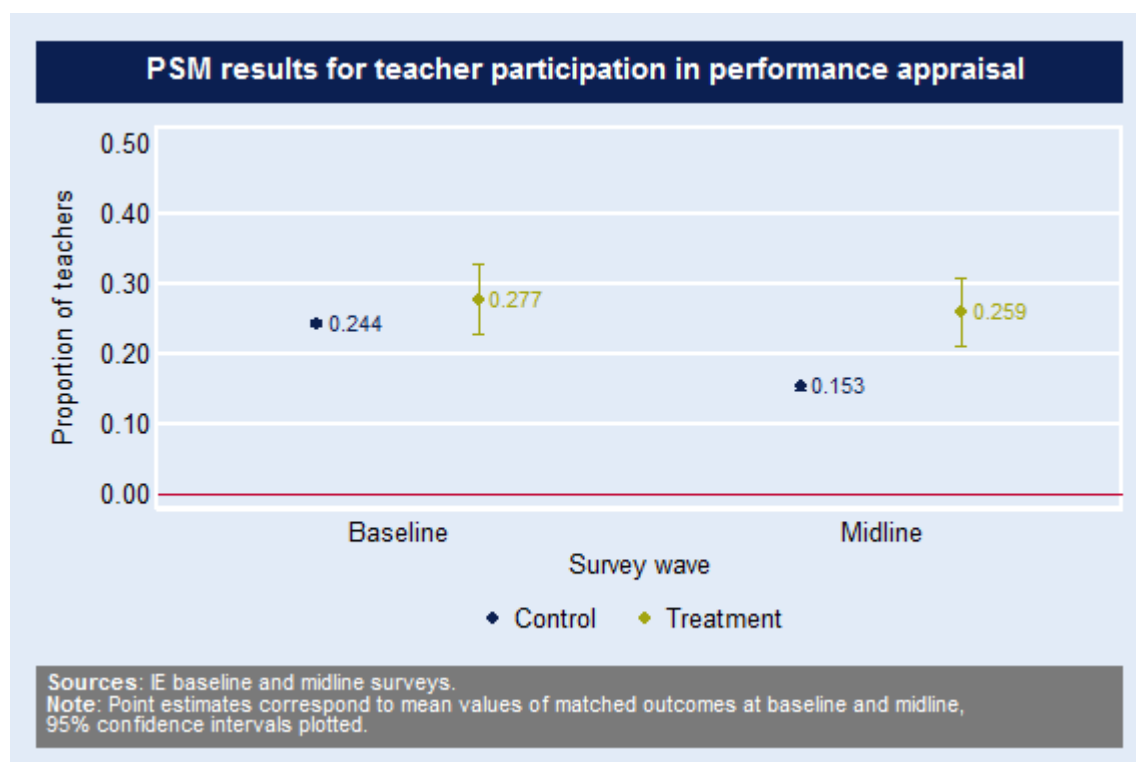
As can be seen below, both the main strategy and the robustness check strategy suggest that teachers in EQUIP-T schools are more likely to report participation in performance appraisals. However, the statistical significance associated with this effect is low on the main model, and therefore, a definitive assessment is not possible.

Table 20: **Teacher performance appraisal: PSM-DID estimate**

	Main strategy	**Robustness check**
PSM-DID estimate	0.073	0.08
P-value (bootstrapping)	(0.08)	(0.04)
P-value (no bootstrapping)	(0.04)	(0.04)

Table 21: **Teacher performance appraisal: Balancing results (Robustness check strategy)**

Balancing results from matching treatment observations across baseline and midline		
Caliper		0.3
N for common support		498
Rubin's B **Rubin's R**	[before matching]	81.08 1.07
Rubin's B Rubin's R	**[after matching]**	**20.61** **1.26**

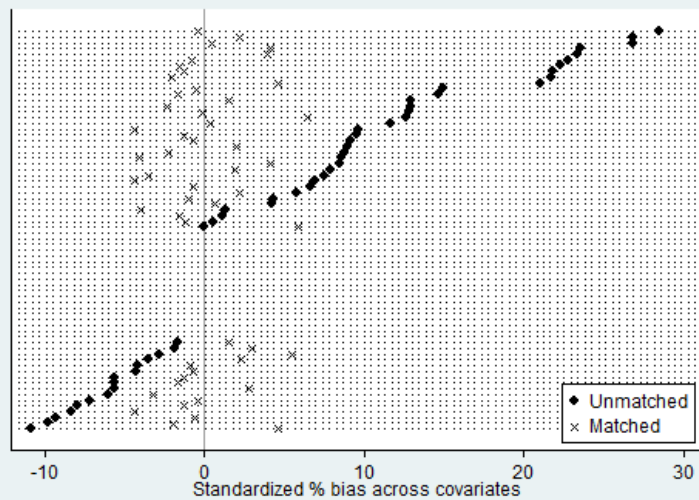# 7   Supplementary descriptive quantitative analysis

This Chapter provides supplementary descriptive trend analysis of indicators of pupil learning, teacher capacity and performance, and SLM in programme treatment schools.

7.1   Pupil Kiswahili and maths raw test score indicators

The tables in this section show trends in raw score indicators of pupil learning achievement in Kiswahili (literacy) skills and maths (numeracy) skills between BL and ML. These results were first presented in the IE ML Preliminary Indicators Report (OPM 2016b), and further details and analysis can be found there. Similar types of raw score-based indicators are reported in the monitoring reports of other large-scale education quality improvement programmes in Tanzania (see RTI, 2016).

Pupil's Kiswahili skills: There have been large improvements in reading speeds, reading and listening comprehension and writing skills. On average, pupils read significantly faster at ML, and this holds across the four different subtests on syllables, familiar words, invented words and a story passage. The size of the change in each case is fairly large. Pupils' reading and listening comprehension have also on average improved significantly since BL, and the gains in average comprehension scores are large. Writing skills have improved significantly as well, with the average scores on spelling and punctuation rising by more than 40% since BL.

Table 22:   Pupils' oral reading speed, BL and ML

Skill area	Indicator	BL estimate	ML estimate	Difference	BL N	ML N
Syllable sounds	Mean # of correct syllables read per minute	20.9	30.4	9.5***	1,491	1,477
Familiar words	Mean # of correct words read per minute	13.7	19.9	6.2***	1,496	1,481
Invented words	Mean # of correct words read per minute	9.3	13.3	4.0***	1,493	1,477
Reading a passage	Mean # of correct words read per minute	21.3	30.0	8.7***	1,496	1,477

Source: IE baseline and midline surveys (pupil Kiswahili test).
Note: (1) Weighted estimates. (2) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 23:   Pupils' reading and listening comprehension skills, BL and ML

Skill area	Indicator	BL estimate	ML estimate	Difference	BL N	ML N
Reading comprehension	Mean test score (%)	19.0	26.8	7.8***	1,496	1,477
	Percentage of pupils who scored more than 80%	1.2	1.7	0.5	1,497	1,477
	Percentage of pupils who scored 0%	55.9	40.8	-15.1***	1,496	1,477
Listening comprehension[3]	Mean test score (%)	31.8	40.9	9.1***	1,496	1,483

Source: IE baseline and midline surveys (pupil Kiswahili test).

Table 24:   Pupils' writing skills, BL and ML

Skill area	Indicator	BL estimate	ML estimate	Difference	BL N	ML N
Spelling	Mean test score (%)	39.1	55.5	16.4***	1,496	1,483
Punctuation	Mean test score (%)	30.0	42.7	12.7***	1,496	1,483

Source: IE baseline and midline surveys (pupil Kiswahili test).
Note: (1) Weighted estimates. (2) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Pupils' maths skills: On the simplest mathematical task on the test, number comparison, pupils' skills have not changed significantly whereas skills on a more complex task, filling missing numbers in sequences, have improved significantly since BL. Pupils' skills in addition and subtraction have on average improved significantly between BL and ML, and this skills improvement applies to pupils with both weaker and stronger mathematical skills. There have also been significant gains in multiplication skills and word problems solving skills since BL. This implies that as well as gains in knowledge of arithmetic operations, pupils' understanding of arithmetic concepts has also improved.

Table 25:   Pupils' skills in number comparison and missing numbers, BL and ML

Skill area	Indicator	BL estimate	ML estimate	Difference	BL N	ML N
Number comparison	Mean test score (%)	64.6	64.4	-0.2	1,495	1,483
Missing numbers in sequences	Mean test score (%)	28.5	33.8	5.3***	1,495	1,483
	Percentage of pupils who scored more than 60%	7.3	9.7	2.4	1,495	1,483
	Percentage of pupils who scored 0%	13.1	7.7	-5.4***	1,495	1,483

Source: IE baseline and midline surveys (pupil mathematics test).
Note: (1) Weighted estimates. (2) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 26:   Pupils' skills in addition and subtraction, BL and ML

Skill area	Indicator	BL estimate	ML estimate	Difference	BL N	ML N
Addition level 1	Mean test score (%)	61.3	68.5	7.2***	1,495	1,483
Addition level 2	Mean test score (%)	30.0	36.3	6.3**	1,495	1,483
Subtraction level 1	Mean test score (%)	45.6	53.9	8.3***	1,495	1,483
Subtraction level 2	Mean test score (%)	19.6	24.4	4.8**	1,495	1,483
Addition and subtraction level 2	Percentage of pupils who scored more than 80%	7.9	12.3	4.4**	1,495	1,483
	Percentage of pupils who scored 0%	37.8	29.6	-8.2**	1,495	1,483

Table 27:   Pupils' skills in multiplication and word problems, BL and ML

Skill area	Indicator	BL estimate	ML estimate	Difference	BL N	ML N
Multiplication	Mean test score (%)	19.4	24.4	5.0***	1,495	1,483
Word problems	Mean test score (%)	28.8	37.3	8.5***	1,495	1,483

Sources: IE baseline and midline surveys (pupil mathematics test).
Note: (1) Weighted estimates. (2) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.

7.2   Teachers capacity and performance

7.2.1   Teacher subject knowledge in Kiswahili and mathematics by topic

Examining subject knowledge by topic (Table 28), teachers of Standards 1-3 perform relatively poorly on grammar and punctuation questions, scoring 44% (42% at BL) on average compared with over 64% on the other topics covered by the TDNA. Teachers perform notably better on direct and indirect tenses scoring on average 72% (69% a BL); the average score on reading comprehension is 69% (68% at BL); and for synonyms and proverbs the average score is 64% (61% at BL).

When it comes to different topics in mathematics, teachers perform the strongest on fractions, decimals and percentages; statistics; and whole numbers, answering on average 69% or more of the questions correctly. The only significant difference between BL and ML is in the average score for fractions, decimals and percentages that is 5 percentage points higher ML. Teachers perform least well on geometry and algebra (average scores 42% and 49% respectively), which are standard six and seven level questions. One possible reason for the relatively poorer scores in algebra is that these questions are located at the end of the TDNA, and some teachers may simply have run out of time for answer these.

Table 28:   Teacher Kiswahili and mathematics subject knowledge by topic, BL (2014) and ML (2016)

Indicator	BL Estimate	ML Estimate	Difference	BL N	ML N
Teachers of Standards 1-3					
Reading comprehension qns correct (%)	68.3	69.2	0.9	247	240
Grammar and punctuation qns correct (%)	42.1	43.7	1.6	247	239
Direct and indirect tenses qns correct (%)	68.6	72.2	3.6	247	240
Synonyms and proverbs qns correct (%)	61.3	64.2	2.9	247	240
Teachers of Standards 1-7					
Whole number qns correct (%)	69.0	69.3	0.3	506	470
Fractions, decimals and percent qns correct (%)	66.7	71.8	5.1***	506	470
Geometry qns correct (%)	43.1	42.3	-0.8	506	470

Statistical qns correct (%)	65.9	71.2	5.3*	506	470
Algebra qns correct (%)	45.4	48.9	3.5	506	470

Sources: IE baseline and midline surveys (TDNA Kiswahili and mathematics).
Note: (1) Weighted estimates. (2) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.

7.2.2   Teacher background characteristics

The personal characteristics, work experience and tenure of teachers of Standards 1-3 have not significantly changed since BL. At ML, 58% of teachers are female and the average age is 38 years, similar to BL levels (Table 29). The share of teachers reaching the official retirement age of 60 years within the next two years is 10%. The average time worked as a teacher is 14 years at ML (16 years at BL) and as a teacher at the current school 8 years.

The highest professional qualifications profile of teachers of Standards 1-3 has not altered significantly since BL. Almost all teachers (96%) have a certificate in education as their highest professional qualification followed by a diploma or advanced diploma (2%).

Table 29:   Teacher of Standards 1-3 background characteristics, BL (2014) and ML (2016)

Indicator	BL Estimate	ML Estimate	Difference	BL N	ML N
Personal characteristics					
Female (% teachers)	55.6	58.4	2.8	327	384
Age (mean years)	39.6	37.5	-2.1	327	384
Work experience & tenure					
Time working as a teacher (mean years)	15.8	13.8	-2.0	327	384
Time teaching at current school (mean years)	8.4	7.8	-0.6	327	384
Near retirement age 60 years (% teachers)	n.a.	10.1	n.a.	n.a.	384
Highest professional qualification					
Bachelors of Education or higher (% teachers)	0.4	0.2	-0.2	326	384
Diploma or advanced diploma (% teachers)	1.5	1.8	0.3	326	384
Certificate in education (% teachers)	94.2	96.0	1.8	326	384
Other professional qualification (% teachers)	3.3	1.3	-2.0	326	384
No professional qualification (% teachers)	0.5	0.7	0.2	326	384

Sources: IE baseline and midline surveys (teacher interview).
Note: (1) Weighted estimates. (2) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (4) Official retirement age is 60 years.

7.3   School leadership and management

7.3.1   HT background characteristics

The average personal characteristics and work experience and tenure of HTs did not change significantly between BL and ML. At ML, 17% of head teachers are female, the average age is 42 years and 10% of HTs will reach the age of 60 within the next two years (Table 30). The average total time worked as a HT is seven years and time worked as a HT at the current school is four years.

There are some significant differences in head teachers' highest professional qualification between BL and ML. The majority of HTs (79%) have a certificate in education at ML, which is significantly smaller than at BL (91%). The second most common professional qualification is a diploma or advanced diploma (17% at ML and 8% at BL). At BL no HTs reported having a Bachelors in Education, at ML this had changed significantly to 4% of HTs. This suggests that at ML, HTs on average have slightly higher professional qualifications.

Table 30:   Head teacher background characteristics, BL (2014) and ML (2016)

Indicator	BL Estimate	ML Estimate	Difference	BL N	ML N
Personal characteristics					
Female (% HTs)	15.8	17.3	1.5	99	99
Age (mean years)	43.5	41.9	-1.6	98	98
Work experience & tenure					
Time working as head teacher (mean years)	n.a.	7.1	n.a.	n.a.	96
Time working as head teacher at current school (mean years)	4.0	4.1	0.1	99	99
Near retirement age 60 years (% HTs)	n.a.	10.0	n.a.	n.a.	98
Highest professional qualification					
Bachelors of Education or higher (% HTs)	0.0	4.4	4.4*	99	99
Diploma or advanced diploma (% HTs)	8.4	16.8	8.4	99	99
Certificate in education (% HTs)	90.5	78.8	-11.7**	99	99
Other professional qualification (% HTs)	1.1	0.0	-1.1	99	99
No professional qualification (% HTs)	0.0	0.0	0.0	99	99

Sources: IE baseline and midline surveys (head teacher interview).
Note: (1) Weighted estimates. (2) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (3) n.a. indicates not applicable.

7.3.2   School characteristics

Average school size and other key resource ratios (for teachers and classrooms) are shown in Figure 21 below. Since BL there has been no significant change in the average number of pupils per school, pupils per teacher, or pupils per class. There has been a small but significant decline in

the average number of pupils per classroom from 78 at BL to 74 at ML (see Annex F, F.3). The overall picture of one of a very constrained physical and human resource environment in most schools. The very wide range of pupil to classroom ratios is also very striking, and the picture was similar at BL.

Figure 21 School size and key resource ratios at ML (2016)
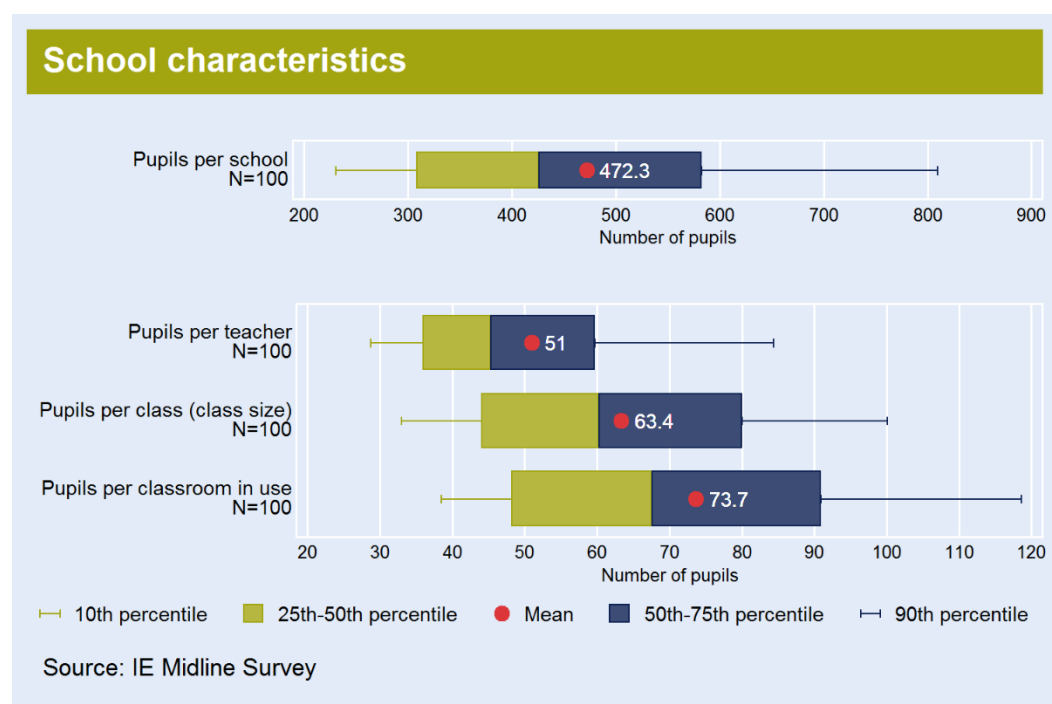
8 Qualitative evidence on teacher capacity, performance and conditions for learning

Improving teacher performance is the core aim of the first component of the EQUIP-T programme. Through in-service teacher training (INSET), provision of tailored teaching and learning materials, and a planned management performance system linked to career development, the programme aims to improve teacher capacity, performance, motivation, and morale, and ultimately aims to improve teaching. As the impact evaluation (IE) baseline findings show low levels of teacher performance in areas such as pedagogy, and low levels of motivation and morale amongst teachers to correlate with high levels of absenteeism, the relevance of component 1 to the overall impact of EQUIP-T on pupil learning is likely high.

In addition to interventions targeted at primary schools (and in particular standards 1 and 2), this component of EQUIP-T established a School Readiness Programme for pre-primary children. The IE baseline found that language was a serious problem for children in primary school, with those not speaking Kiswahili at home performing significantly worse in both Kiswahili and mathematics. The SRP is expected to help children with a different mother tongue prepare for learning in Kiswahili.

The findings below assess the expectations and perceptions of change in this programme component, based on the qualitative findings from EQUIP staff, education managers including WECs, and stakeholders from nine case schools across three districts. In each district, the research team visited the same schools as at baseline, sampled to be one low-performing, one typically performing, and one high-performing school.

8.1 Teacher training

Teachers in all nine case schools have attended EQUIP-T INSET training, and respondents feel training has a positive effect on teaching. Teachers state training to make them feel more motivated to teach, as *"…one may have been to the college for a long time and has forgotten many things, but when you go to the trainings you meet with other teachers [from] different schools"* (Teacher, School 3, District A). Respondents describe INSET as providing teachers with further skills, as well as functioning as a 'refresher course', allowing teachers to feel more confident in their teaching and thus more motivated.

Schools and senior education managers perceive EQUIP-T to be responsible for the decision to introduce the 3Rs curriculum, and highlight this as the main contribution of the programme. Teachers feel focusing on three subjects at a time allows them to develop deeper subject knowledge within specialised subjects, instead of dividing their focus across several subjects. EQUIP staff noted that the concurrent introduction of a new syllabus has contributed to changes in classrooms, with teachers finding it easier to manage fewer subjects and concentrate on these skills. **Additionally, teachers consider teaching methods and the knowledge on how to prepare and use teaching aids as main takeaways from INSET.** The ability to break down lessons and subjects into smaller portions that pupils are better able to understand makes it easier for teachers to teach, and respondents' state training to have helped them understand the importance of lesson preparation. A teacher thus noted:

> *EQUIP-T activities have helped us a lot, through confidence, preparation [and] new skills on teaching. Therefore, the way we access new teaching skills, lesson preparation [and] making resources helps one to have confidence, which is quite different from other years* (School 3, District A).

As such, teachers feel more confident about preparing for and teaching classes. Though some teachers refer to *"understanding subject matter"* (Teacher, school 1, district C) as a main reason for increased confidence, the majority of case schools reference pedagogical techniques (for

example child-centred learning, group work and dividing subjects into smaller parts) and preparation of low- and no cost materials as having affected their confidence.

8.1.1   Logistics of training

Teachers in FGDs feel that the time allocated to INSET is inadequate, and that additional days are needed in order to effectively cover all modules. The majority of case schools report that training sessions try to cover too much content in too little time, with teachers finding it difficult to understand all the new information. For instance, a teacher noted:

> *"We go for training, but the challenge is time. In two to three days…we learn lots of things, but not all will be well understood. Some of them you may not understand… Enough time should be planned, and the topics should be taught [long] enough, because not all people learn quickly. There are [fast] and slow learners"*[15]   (School 3, district A).

Consequently, teachers in all case schools feel they are not able to learn as much as they potentially could, due to an overload of information. Though adding more days to INSET may prove logistically difficult,[16] the effectiveness of training may increase as teachers are able to absorb and reflect on training modules better if they are spread out. For example, as one teacher explains: *"those four modules for arithmetic you cannot learn five modules in a day. I propose each module to have its own day"* (school 2, district C). As such the division of modules across time may further depend on the content of each module, with teachers stating initial new content such as learning about the new curriculum to be more difficult than learning new teaching techniques.

Teachers in all schools visited highlight allowances as one of the main benefits to attending training. According to a head teacher:

> *"The attitude is good towards the trainings. However, we have been dominated by the culture that is not that good in which the teachers have maintained that if you go for training, you have to be paid…in [the] real sense if you go for training I have to pay myself, if you go for school you must pay the fee, but teachers do not encounter that [as] part of their work"* (School 2, district B).

Though teachers in FGDs acknowledge the benefits they gain from training in terms of increased confidence and motivation, the emphasis is still on allowances as the primary attraction amongst the majority of teachers. As exemplified by the above quotation, some teachers feel that they are attending training as a favour to the education system, rather than as a benefit in itself through the positive effect on their professional development. Teachers refer to training sessions as being challenging, with long days, and so they see allowances as somewhat compensating for that.

> *"Teachers are carrying more burdens in teaching 31 periods and after the class hours they have to attend EQUIP-T training, oh gosh! They [are] really tired"* (Teachers, school 1, district C).

In this manner, the majority of teachers seem to value the allowance received from training above the potential knowledge gained. A head teacher (school 2, district B), refers to the 'aid' system behind training programmes, and that since teachers have never had to pay for additional professional development, the view has formed that one should not ever be expected to pay. Respondents in several schools highlight this as a potential issue with regards to the sustainability of training, as teachers may be less likely to attend INSET were allowances to disappear, or even

---

[15] Some quotations have been corrected grammatically whilst carefully considering the meaning behind a quotation. This is since grammatical errors are likely to have occurred as part of the translation process rather than reflecting the respondents grasp of the language.

[16] A longer INSET could be more expensive, time consuming, and difficult to manage because of a higher turnover during the extended training period, so we acknowledge that extending the length of the training could pose a different set of problems.

diminish. Thus, though teachers display an overall positive attitude towards INSET, this appears highly related to receiving incentives.

8.1.2   The cascade and in-school models

An INSET coordinator is responsible for coordinating how to deliver modules for in-school training, as well as supervising the teachers in class. Schools select the INSET coordinator based on whom the head teacher and other teachers believe able to deliver the content of training to others (head teacher, school 2, district B). This means that the INSET coordinator is not a teacher in standard 1, 2, or 3 in all cases, as schools frequently appear to choose the current academic master/teacher to also take on the role of INSET coordinator.

Teachers say that the training they get is not sufficient to allow them to share that knowledge with other teachers. In all case schools, teachers express challenges in conveying the content to other teachers, *"…everyone has a different level of understanding in the training so, a person may not understand well so that he can deliver to other teachers in school"* (Teacher, school 3, district B). In this manner, respondents state that the materials covered, such as phonology or arithmetic, are something that teachers themselves are still learning when attending training and not something they are able to effectively teach other teachers how to teach. Teachers therefore find it challenging, as whilst they are able to reiterate what they have learnt, they struggle to explain it properly:

> *"There is a difference because the way [the] trainer facilitated will be different from the person who received training so that he or she can train others. Even when asking questions the person who received training to train others will not be able to answer"* (Teacher, school 2, district C).

Some teachers refer to picking up certain aspects of training to share with other teachers in school, such as how to teach consonants, rather than training them on the wider curriculum. Knowledge sharing thus occurs selectively based on what teachers feel they themselves have picked up. **As such, respondents state the knowledge gained from training varies significantly between those teachers who receive INSET from District INSET Teams and those who receive INSET in the school.**[17] Peer-to-peer training consequently will be highly dependent on the level of understanding of the teachers that attend, and how successfully it is implemented is likely to vary significantly from school to school. Importantly, education staff interviewed say they have asked EQUIP to reduce the number of tiers in the INSET cascade model, to at least make sure that teachers from all schools have direct contact with professional tutors. Both case school level and district and regional levels thus highlight this issue in design.

Additionally, schools find it difficult to organise training in the school, both due to logistical and motivational challenges. This relates to the notion of allowances as motivating teachers to attend INSET, with in-school teachers perceiving the system as unjust since only those attending district/ward level training receive allowances. Consequently, teachers find it difficult to organise in-school training as their peers are unwilling to attend without receiving some allowance.[18] As teachers in school 2, district B discussed:

> *"The trained teacher is supposed to teach other teachers in their respective schools but it's hard to do so because other teachers refuse to attend the class saying that 'you attended the training and get money while we just stay here, listen to you and get nothing'. It's really challenging".*

---

[17] Professional tutors from the teacher training colleges deliver INSET at district or ward-cluster level.

[18] It is important to note that the qualitative research only sampled teachers who have attended district-run INSET. The analysis and evidence around motivational challenges thus need to be taken as perceptions by those teachers spoken to and cannot be fully triangulated or understood.

On the contrary, education managers were of the view that once more senior staff explained the importance of training to teachers, they were accepting of the lack of allowances. For instance, according to a WEC:

"There were challenges at first, as many teachers were requesting payments as other teachers who had attended the same training organised by EQUIP were paid. After we told them importance of the training, it was well received by the teachers, it is now calm." (WEC Y, district C)

Relatedly, schools organise school-based INSET at the end of the school day, with teachers reporting being too hungry and tired to focus. According to teachers in school 1, district C:

"The difference is that you may find that those who attended training elsewhere have allowances but when in school you just sit there without even having something to eat. This is challenging to some teachers who are living in the neighbouring village who stay the whole day without having something to eat. It is also challenging to coordinators to provide seminars to others. They know that you get allowance from the same training. At the same time you tell them to stay for an hour after 2pm for seminar but they will see it as something meaningless to them"

Thus it appears that lack of food at training is not only a motivational issue but also a logistical issue, since keeping teachers after school for too long without food is not possible. As such, this directly affects the amount of time spent on in-school training, with seminars kept to under an hour in most schools visited, when they do occur. This may thus further affect the difference in the effectiveness between district-run and school-based training.

The programme thus needs to consider strategies to reduce the risk of teachers not attending school-based INSET due to perceived lack of incentives, and it is worth considering whether the programme can provide food for teachers during in-school training. Additionally, the pace of training needs to be further explored with regards to the amount of content covered per training session. It is important to assess the trade-off between teachers attending fewer sessions (potentially too content-heavy for teachers to learn/process) or more sessions (that may be costly and keep teachers away from the classroom more frequently). The balance between cost and length of training requires further investigation to ensure the most effective model.

8.2   Teaching and Learning materials

All schools have received materials from EQUIP-T, with respondents in particular highlighting manila papers and marker pens. All schools, including parents, teachers and pupils, further state that there is a visible change as teachers now make teaching aids and use them to decorate the classrooms. The research team observed manila papers in all schools visited as part of the qualitative sample, however, this often meant one or two posters simply displaying the alphabet or the numbers 1-10. The extent to which these are used any differently from how teachers already use the black board is questionable. The qualitative research team did not see teachers use any other teaching aids, though pupils arrived at schools with home-made counting aids for maths.[19] **Teachers state that receiving teaching materials motivates them, and helps them simplify teaching activities.**

With regards to textbooks, respondents in all schools refer to not having received curriculum specific textbooks, which hinders effective teaching of the new curriculum. There is some perception amongst district level respondents that that the EQUIP training has mitigated this problem: *"from EQUIP training we are capable of teaching by using previous books"* (WEC Y, district C). Still, teachers in all schools say they face difficulties using books that

---

[19] The qualitative research team were not carrying out structured lesson observations, but made time to informally observe lessons and break times.

specifically cover a subject rather than 'how to teach a subject', hence facing challenges translating knowledge on which pedagogical techniques to use when teaching a subject, and to prefer curriculum specific textbooks that link techniques with content. According to a school teacher:

"There are changes in grade one and two and their curriculum is different from other classes. So, I would like to get books because I don't have books for teaching new curriculum. I just use books of the former curriculum to teach pupils. Those books are quite different, there are no directives and pictures. So, I have to use my own knowledge (School 2, district C).

Instead, teachers refer to using 'publications' on the syllabus and curriculum as a reference in teaching, and **the majority of teachers feel they would not be able to teach the new curriculum if it were not for these guides**. However, an additional challenge is that the guides refer to teaching practices correlating with specific teaching aids that schools do not have access to, though the qualitative findings were not able to tease out which specific teaching aids this referenced.

In terms of additional materials, schools interestingly reported marker pens, glue, and tape as highly in demand, and that they are not able to produce more teaching aids using manila papers as they have run out of these basic materials. A head teacher further noted:

"We also have deficit of manila to make teaching aid. When the teacher goes to teach a class he should have manila to make teaching aid so you find that a teacher has one manila for all class and he needs to make drawings which needs more manila. The rest are lack of exercise books to make lesson plan (School 1, district B).

Teachers in all schools thus requested additional materials, in particular manila papers and marker pens, so that they can continue to make and update teaching aids. The environment where the schools are located mean that humidity and wind easily ruins papers. The access to teaching and learning materials needs to be sustainable. EQUIP may need to continue to supply these materials, or alternatively encourage school to purchase these materials with the money they receive from EQUIP (when schools are located in areas where they are able to access these).

Thus, schools in all districts refer to an overall scarcity of resources, in particular textbooks. Teachers comment that *"those books we have right now are torn out and very old…we didn't receive books for years…That [is] why you may find a class has one book. In this case, I'm not happy with my teaching"* (school 1, district C). Schools report not having sufficient books for all pupils. Pupils themselves say they borrow books from the school and then share. For example, in school 1, district A, pupils would borrow books for different subjects for a year, and then they would swap those amongst themselves in groups of three when they needed to work on that particular subject. However, some of the schools visited had next to no access to textbooks, with the teacher copying the textbook onto the blackboard.

8.3   Teacher performance: subject and curriculum knowledge

Schools visited perceive the quality of teaching to have increased, mainly because of the new curriculum. Respondents in all schools (including head teachers, teachers, community leaders, SCs and parents) feel that teaching has improved, and that children are now able to read better than before. Parents also highlight that pupils are now clear on when they have each subject and therefore when which homework has to be completed. Teachers similarly feel more confident focusing on 3Rs, and say that even though there are challenges in learning a new curriculum, training helps them understand how to teach these subjects well rather than trying to teach many subjects less well. The baseline findings showed that due to the lack of teacher training, teachers were unaware of the content of the new syllabus and did not know what subject matter should be taught to which grade level pupil. Teachers now appear to be knowledgeable about what is

included in the new curriculum and clearly emphasise having understood the need to teach standard 1 and 2 pupils the 3Rs before any other subjects, as noted by a head teacher:

*"EQUIP has seen the importance of 3Rs because children were studying many subject compared to their ages. Children's ages were small, so it was not easy for teachers to teach them about six to seven subjects...therefore these teachers were happy on the fact that they will teach three subjects" (*School 1, district A).

Moreover, teachers are now preparing lesson plans prior to teaching. Respondents in each school comment on lesson plans as an important reason for why quality of teaching has improved. Teachers in FGDs consider lesson plans as helping them feel confident in class, and head teachers in three of the schools state that it also helps them (or the academic master) to better be able to monitor and help teachers. Preparing lesson plans mean that teachers are aware of what resources they will be needing during a particular lesson, and as such that they can show up to class with a clear idea of how a lesson will run.

"[Interviewer: in case the observer came into your classroom today, what changes would they notice compared to two years ago?]

"Teacher: Previously, you could have fear of what you are going to deliver due to the fact that we were doing things based on experience. Currently, we are confident because I prepared the lesson plan in relation to the training we received and I can be able to make teaching materials according to the topic as well as how to use the materials" (School 2, district B).

Additionally, respondents perceive subject knowledge to have increased. Through the focus on 3Rs, teachers are able to focus on three subjects in depth, and schools consider this the main reason why teacher capacity has improved. At baseline, the qualitative research found that teachers felt confident in their knowledge of the subject in which they specialised, but that many teachers were teaching subject areas in which they had not specialised. The mid line qualitative research found that teachers are able to have a more targeted focus in how they prepare for classes through the 3Rs, and they refer to this 'new system' as helping them feel confident in their knowledge of the subjects they are now teaching.

District and Regional interviews found that EQUIP training was seen as the first targeted training for early grades in many years. Senior managers explained that early grade teaching was a major part of the syllabus in teacher training colleges in the 1990s but has since reduced, meaning that recent graduates have not received relevant training for the lower standards. As REO E states: *"We don't have so many competent teachers for the early grades. We had some time ago, but after stopping some of the programmes that we had, it reached us into getting teachers who were really incompetent in teaching early grade, especially standards 1 and 2. Most who had high experience in early grades have retired."* Education managers expressed gratitude to EQUIP for coming to build capacity for these teachers.

Schools consider teacher's subject knowledge to have improved as INSET training has helped locate and plug gaps in teacher's subject knowledge. One particular example brought up by teachers, head teachers, and managers is the focus on pronunciation.

Teachers and managers feel that the teachers are better able to teach reading and writing, both through themselves increasing their knowledge around pronunciation, and through learning how to teach by breaking each word down into syllables. A head teacher noted:

"...previously we used to teach sounds in these classes in a way that affected the reading ability...they were only taught letters e.g. ABCD...those are the names of the letters but not the sound of the letters. So sounds were not taught completely but now they are teaching

sounds, which makes it easy for the children to understand how the letter sounds…" (School 2, district B).

All schools cite this as a main effect of the INSET training that they have implemented effectively. This way of teaching reading was further observed by the qualitative research team whilst in the schools, both during classes and when observing children sitting outside reading or doing their homework.

8.4   Teacher performance: pedagogy

Compared to baseline, the majority of teachers now identify pedagogy as a knowledge required of a teacher, highlighting teaching methods over subject and curriculum knowledge. Teachers and managers also discussed child-centred teaching and participatory methods, though it is not clear from the qualitative findings to which extent teachers have a deeper knowledge of what these methods mean or when to use a particular method.

Additionally, teachers state a change in their own attitude towards pupil learning, emphasising how INSET training has helped them understand how to teach standard 1 and 2 pupils so that they are able to follow the material. Whilst at baseline teachers expressed frustration that they could not understand why pupils in their classrooms could not grasp certain topics or display certain skills, the qualitative research found teachers to express their new found knowledge around 'how to teach' as a main takeaway from EQUIP-T training. As one teacher mentioned when discussing INSET takeaways:

"…I was thinking that it is hard to teach grade one and two. In reality it is not hard, because it depends on how you break down the topic into small sections, and to make the pupils know what is learning" (Teachers, school 2, district C).

Though the qualitative research did not include a structured qualitative classroom observation (QCO), the team carried out non-participatory observations of classes in each of the schools, triangulating this with answers from FGDs and KIIs and situating responses. In the majority of schools visited, the research team observed teachers beginning a lesson by asking questions around a prior lesson, and writing on the board what they would focus on in that lesson. The quantitative data from baseline shows that only 22% of teachers checked if pupils had acquired new skills or knowledge as part of the lesson introduction. FGDs with pupils further emphasise this change, with pupils stating that teachers ask them questions about lessons and introduce a new topic, elaborate on it, and then let pupils work independently, *"they give an example, elaborate it, then they rub the example and we start writing"* (Children, school 3, district A).

The qualitative research team observed teachers using participatory methods, as well as using teaching resources such as the blackboard, manila paper, and marker pens. The use of child-centred learning activities were highlighted in FGDs and KIIs in all schools, with several parents referencing 'more creative' methods to be used by teachers now as compared to previous years, *"…you know young children like games, so if you put for them some games and songs, the brain is tuning itself in a good way"* (Mothers, school 1, district C). Overall, respondents in all schools feel that INSET training has taught teachers to make learning more fun for pupils.

"[Interviewer: if an observer came into your classroom today, what changes would they notice compared to two years ago?] Teaching techniques and the level of understanding among pupils, teaching aids that you make, participatory methods of teaching such as group work, individual or in pairs, the use of drawings, pictures and publications" (Teachers, school 1, district C).

Schools perceive this notion of using a variety of methods as an indication of increased capacity of teachers. However, it is less clear from the findings the extent to which teachers know when to

make use of each method. For instance, teachers speak of involving pupils more when teaching the class, and state *"...they have to be involved as well as become key participants in the class, and we just give them directives and see what they are doing"* (Teachers, school 2, district C). In some schools, pupils referred to teachers letting them teach each other subjects, or to get directives and then work independently. The extent to which teachers were actively involved in the lesson seemed to vary between schools and in some cases the qualitative team observed teachers simply writing an activity on the blackboard and then leaving the pupils to it, and then focusing instead on marking. When talking about changes in the classroom in the last two years, a teacher noted: *They will find three quarter of the pupils in the class are doing things on their own and you just direct them"* (School 1, district B).

In a class in another school, however, the teacher provided assignments for pupils to do in the class, and would pass through the class and mark for each pupil. Afterwards she would involve pupils in doing corrections on the blackboard (school 2, district A).

Corporal punishment appears prominent in all case schools, in spite of Head Teachers and teachers stating to have learnt alternative methods of class control through EQUIP-T. Teachers and Head Teachers often speak of learning the importance of befriending a child, and not being seen as 'the enemy', as this quote from a teacher demonstrates: *"[if you] use harsh language the pupils will be afraid of that and even if they have something to say they will not speak out"* (School 2, district C).

At the same time, pupils in all schools consider caning the main thing they dislike about school. Pupils noted that teachers use canes or *"beating us with sticks"* (school 2, district B) to make them 'understand'. It was also clear from qualitative observations that canes were being used. Yet, head teachers referred to the decrease of corporal punishment as a change caused by EQUIP-T. For example, the head teacher in school 1, district C explained:

> *"They (EQUIP) were discouraging on corporal punishment and using sticks to the students. They said there are some of the things which can be used in class without using of sticks and will make students understand the class."*).

As conversations with pupils and informal observations highlight, however, though corporal punishment may have decreased, it is still actively used as a method of controlling the class. The discussions with teachers and head teachers are likely to display an element of social desirability and thus response bias, as those who have received INSET will be clear that EQUIP-T discourages corporal punishment, making teachers and head teachers likely to highlight this as a change regardless of the extent or significance of this change. The pupils' accounts thus provide a strong check-point for the strength of this evidence. One possible focus for the qualitative data at end line is to explore the extent to which schools have actively implemented a child-friendly learning environment. EQUIP-T might need to strengthen components on corporal punishment to increase teachers' awareness, and in particular to strengthen their classroom management skills and equip them with strategies other than corporal punishment.

8.4.1 Inclusive instruction

Schools appear to make use of peer-to-peer learning as a way of ensuring inclusive instruction. Respondents in all schools, including parents and pupils, state that pupils work actively together so that 'slow learners' can learn from 'fast learners'. As one mother noted:

> *"…my grade three children said 'mother, I know how to read and I am given someone to assist her in reading'…This is the technique that the teacher uses by asking those who know how to read to assist others who do not know how to read"* (School 3, district A).

Several teachers explain that they group pupils together to learn from each other, and to have

learnt this focus on including 'slow learners' from EQUIP-T. Rather than considering pupils as 'less intelligent' and unable to learn, EQUIP-T has highlighted that pupils may just be slow learners, or be affected by problems in their homes, and thus teachers explain having understood the importance of actively engaging with these pupils.

Yet, though teachers and Head Teachers perceive INSET to train them well on how to include children better, teachers experience challenges implementing these methods in the classroom. Whilst teachers may be better able to identify students needing extra attention, the large class sizes are cited as a reason why they are unable to give that extra time. According to a head teacher, *"…the classes are overcrowded…in a class of 100 students, how can you use 30 minutes for one hundred-and-something people? How can I identify pupils with different needs? It is very difficult. You can but for very few, and the significant challenge is the number of pupils"* (School 2, district B). Teachers thus face difficulties in spending extra time on some pupils, as the class size already means that pupils in general are not getting sufficient help. **All schools identify over-crowdedness of classrooms as a key challenge for inclusive instruction.**

Five schools visited report teachers to put on additional instructional time for what they describe as 'slow learners'. Teachers in these schools put pupils into groups depending on how quickly they grasp a concept, and then either ask some pupils to stay behind after class or attend extra classes on the weekends. According to a head teacher, *"…there are remedial classes for learning 3Rs for slow learners in order to cope with others so a child can attend, and we have selected few teachers to help them after the class hours"* (School 3, district A). The qualitative team observed teachers in two schools who would remain with a pupil after class finished to go over concepts the pupil had not been able to grasp.

This notion of additional instructional time for 'slow learners' is further emphasised by parents as well as school committees in these five schools as reasons for why pupils are now better at reading and writing. It is also used to reflect upon teacher's motivation,[20] as an example for how teachers are 'trying harder' than in previous years. The five schools that reported additional instructional time in this manner all had at least some housing available for teachers near the schools. In the four schools where additional instructional time was not reported, it was still referred to as needed, and the reason for why teachers did not provide this was due to lack of teacher houses.

However, even in these five schools, teachers say they find it challenging to get pupils to attend additional class time, as parents are not willing to let them partake in school activities outside of official school hours. Teachers say that parents' attitudes towards education, with a low awareness of the benefit of education, affects their willingness to allow children to attend extra classes. Parents and teachers both link this to opportunity cost, with parents needing children at home to do work such as farming or looking after cattle. As such, parents feel that they are already 'allowing' children to spend a significant amount of time in school. A teacher reported that parents often complain to them, saying *"…why are you asking children to come to school on a Saturday while you were there with them from Monday to Friday? When should we use them?"* (School 2, district B). This may also relate to the distance to school many pupils experience, as keeping pupils after school closes mean they would only return home in the evening. Additionally, with all schools visited unable to provide children with lunch at school, after-school classes may be affected by pupils being hungry, further complicating the logistics of additional instructional time and limiting it to weekends, when, as discussed above, parents may be unwilling to send their children to school.

8.4.1.1 Gender balance in teacher interaction with pupils

Schools in all districts perceive gender-balance to have improved. Head teachers and teachers perceive EQUIP-T to have improved gender-balance in two ways, 1) through helping

---

[20] Discussed further in section 8.6 on teacher motivation and morale.

build confidence amongst girls in learning and 2) for teachers to now know the importance of involving both sexes. Both boys and girls in FGDs said they are able to answer questions in class, and the qualitative team observed teachers using gender-sensitive techniques such as alternating between boys and girls when asking questions in the classroom. Teachers link this to having increased girls confidence, as *"…selecting two girls and two boys to go before the class…is quite different from some years ago that only the one that had shown her/his hand was the one to demonstrate…"* (Teacher, school 1, district A). Thus, previously teachers would only include those children who themselves raised their hands, and as girls tended to be less likely to do so, boys ended up more actively included. However, it is important to be cautious exploring this evidence, since, as with questions around corporal punishment, there might be an element of social desirability bias affecting teachers' answers and class behaviour. Moreover, through gender-sensitive techniques such as alternating between boys and girls are useful, it is not clear whether there is an intrinsic change in gender attitudes amongst teachers and communities. Girls in all schools appear more likely to do chores during class times (including girls in lower standards), and the team observed girls looking after teachers' babies and cooking lunch in teachers' houses. Girls consequently lose out on instructional time. The qualitative team observed this in all schools visited.

There are gender considerations with regards to out-of-school responsibilities for both boys and girls. Whilst girls appear to spend more time on household chores such as cooking and childcare, boys seem to be actively involved in income generating activities (IGAs) like farming or looking after cattle. In pastoralist communities in particular, it appears that boys may have less time than girls to spend on school work, as boys need to be away with the cattle, and are thus likely to miss school days. Lastly, early marriage is highly prominent in all schools visited, and though this does not appear to directly affect attendance of girls, it has the potential negative effect of affecting parents' attitudes towards the need for girls' education, and girls' own attitudes towards why they should study – though this may be more likely to manifest itself in later school years. Changes to deeper entrenched gender roles need further exploration to better understand any variance in performance between boys and girls.

The over-crowdedness of classrooms and a lack of desks for all may have an adverse effect on inclusive instruction. The qualitative baseline found most teachers to show gender balance in classroom interactions with pupils, but to interact more with pupils at the front of the classroom than in other parts. The qualitative midline did not observe this spatial difference nor was it brought up by respondents, which is supported by the quantitative findings of improved spatial balance. All schools referred to having received additional desks, either through the government or from community contributions. Teachers report desks to still not be enough, and for children to often have to sit on the floor. Moreover, with the classrooms overcrowded, a head teacher noted that teachers cannot move around to assist everyone in the classroom. In particular, where pupils sit on the floor, respondents say teachers are less likely to mark their work, or take the time to help them during class. Teachers themselves acknowledge this, and feel it is difficult for them to attend to those seated on the floor. There is thus a question whether spatial balance is the best indicator of better inclusion or whether this relates to teachers focusing on those pupils seated on desks, whether in the front or back of the classroom. According to a community leader:

"It was one month back when I received a complaint from one of the parents that the exercise book for her child was not marked. When the parent was trying to ask her son the son said 'we who are sitting down (on the floor), teachers do not mark our exercises. They mark exercises for those who sit on the desk'. When the parent went to school to complain about this, teachers admitted to not marking the exercise for these children" (School 3, district B).

In this manner, there is a potential link with a lack of gender balance. The qualitative team found through classroom observations and informal conversations with girls that, in at least three of the schools visited, girls are less likely to sit on desks than boys. Girls in school 1, district A say that if they sit on desks when they arrive, boys come and push them off and make them sit on the floor. Though the current evidence does not fully explore this, it is important to be cautious of the

potential effect this may have on gender balance, as well as for including pupils (boys and girls) left without desks.

8.4.2   Instructional time

Some teachers perceive instructional time to be too short to teach the new curriculum.
As a teacher notes:

> *"The time is not enough during this time of teaching. For instance, to use thirty minutes of teaching until the students understand…you can use sixty minutes to teach one subject for a child to understand but thirty is not enough" (*School 2, district A).

This is not reported as a problem in all schools, and the qualitative team observed many classes which only lasted for around 20 minutes. This was the case in classes that qualitative researchers observed officially as well as in other classes taking place on the days when the qualitative teams were visiting. It is possible that teachers are not fully confident in planning lessons based on the available time, but rather based on how they would approach teaching a topic fully, and so lesson plans may have to carry into the next class as well. Additionally, there may be a link with the large class sizes, as taking the time to teach so many pupils at a time would affect the time it takes to effectively teach a subject. This relates to teachers' comments that INSET training does not fully consider the reality of the teaching context, and for teachers to find it difficult to implement what they learn within their schools.

Teacher absenteeism[21] appears to influence instructional time, but schools perceive this to have decreased due to the president's slogan *hapa kazi tu*[22]**.** In all case schools, teachers appear to be present in school, though not necessarily in the classroom. However, all schools claim teacher absenteeism has decreased and that teachers now work harder. The main reason given is *hapa kazi tu,* the president's election campaign slogan now associated with the new government, which encourages all to work hard at their job or they will be 'punished/replaced', and if they do work hard they will be 'rewarded'. This notion of national change in work ethic is highly stressed by all respondents as affecting teacher attendance and commitment. However, it is important to consider whether this will lead to a lasting effect or whether the perception of monitoring will wear off, especially if no punishment/reward materialises.[23]

All teachers in FGDs complain that their workload is too high, with too many classes to teach, affecting their ability to attend all classes. Teachers themselves consider timetabling issues, where teachers have too many periods to cover, to prevent them from attending classes or having to cut class time short to allow time for marking and lesson preparation, as the quote below shows:

> *"When you see the timetable I start teaching from 8 to the end of the period. That is, you get out from one class [and] into another class…So, there is no time for you to get [a] free period because sometimes you may have so many exercise books to mark and you have to do it while missing other periods"* (Teacher, school 2, district C).

Teachers appear to spend a significant amount of class time marking pupil assignments, either inside or outside the classroom.[24] Teachers mention 'marking' as the main reason for not being able to teach. Other respondents' accounts, support this; pupils say that teachers at times leave them alone in the classroom to go to the office and 'mark'. The qualitative team also

---

[21] Discussed at length in Section 8.6.1 (teacher absenteeism)

[22] Online discussion translates this as 'Just work' or 'Here it's only work'.
https://kabatilakanga.wordpress.com/2015/11/20/hapa-kazi-tu/ and http://thisisafrica.me/no-christmas-tanzania-hapa-kazi-tu/

[23] This notion of perceived 'monitoring' is discussed further in Section 8.6.2 on monitoring and accountability.

[24] Discussed further in Section 8.4 (teaching activities).

observed teachers marking assignments during class hours in three of the schools. Interestingly, marking is also highlighted in discussions around 'monitoring and accountability'.[25] Respondents cite marking of exercise books as a way for parents to keep track of what their pupils are doing in school as well as the means through which inspectors and WECs check teacher attendance. Combining growing enrolment with increased monitoring of teachers through pupil exercise books may thus result in teachers spending additional time marking assignments, making a trade-off between marking and instructional time. The quantitative findings mirror this notion of 'workload' as a reason for lost instructional time. Similarly, head teachers in all schools state their administrative duties affect the time they have to spend teaching, with head teachers facing a trade-off between school management and teaching responsibilities. At end line a priority for qualitative research could be to explore further whether or not there is a relationship between monitoring and instructional time.

Lack of classrooms mean that standards switch, with Standard 1 attending in the morning and Standard 2 in the afternoon. This means that standards do not receive a full day of teaching, but instead their potential instructional time ends up getting reduced. However, it is important to consider whether longer days for pupils would increase learning, since schools, as mentioned above, are unable to provide food for pupils, which leaves pupils hungry and unable to concentrate. Yet, there is an infrastructural challenge where classrooms end up getting shared between standards. Several schools have addressed this by sharing classrooms, where two standards use a classroom at the same time. In one school visited, standard 2 shared a classroom with standard 5, facing different walls, and teachers took turns teaching. This does not only reduce the lesson time in half, but may also affect learning for particularly the lower standard (in this case standard 2). The qualitative team observed standard 5 students answering questions posed to the standard 2 pupils. Moreover, sharing a classroom with older students may affect the confidence of the younger pupils to speak up and actively participate in class. This loss of instructional time is acknowledged by schools to directly affect learning, *"…if you check for the classes that are taught all the periods and those that are not taught all the periods you find that when you give them exams in classes that are taught all subjects there has been improvements in academic performance has fallen"* (Head Teacher, school 2, district B). All schools visited are in need of additional classrooms.

Moreover, the increased enrolment and addition of a pre-primary class adversely affects instructional time for standard 1 and/or standard 2 in some of the schools visited. With the introduction of compulsory pre-primary at each school,[26] respondents perceive enrolment to have increased. This is in line with the quantitative findings. As schools need to introduce yet another standard (pre-primary), the challenge of lack of classrooms intensifies, and in the cases where schools do not have sufficient numbers of classrooms, they either teach pre-primary students outside, let them share a classroom with standard 1, or group standard 1 and standard 2 together. In cases where pre-primary students are taught outside, this ends up highly weather-dependent and the qualitative team observed pre-primary classes having to move every 20 minutes in order to stay out of the sun, disrupting the lesson.

Teachers state pupil absenteeism as a main reason for loss of instructional time. When pupils are away from class, they do not only miss that particular instructional time, but teachers indicate it to be difficult to allow sufficient time for these pupils to catch-up when they do attend, resulting in the absenteeism of some pupils, indirectly affecting the instructional time of other pupils. Additionally, as discussed in Volume I Chapter 2 on pupil learning, respondents' state pupils from pastoralist communities have higher rates, and longer periods, of absenteeism. Respondents in schools located in areas where pastoralism is prominent noted this consistently.

Distance to school also appears to influence instructional time, with teachers and pupils living far away needing long lunch breaks to have time to go home and eat. The inability of

---

[25] See section Section 8.6.2 (monitoring and accountability)

[26] As stipulated in the second Education and Training Policy (ETP II), 2014.

schools to provide food at school directly affects both instructional time and learning. All schools but one have a long break during school hours to allow teachers and pupils to go home to eat at lunch. Yet, pupils who live far away rarely go, and they end up spending a full day in school on an empty stomach, with the longer lunch break meaning they have to remain hungry for longer in school. Additionally, both pupils and teachers reported that they go home to eat and then not come back, or come back late, at times, decreasing instructional time. In one school they have decided to cut breaks short and instead close earlier to try to combat this issue. All schools fondly refer to the World Food Programme (WFP), as they had previously provided food in schools. This programme was seen to have had a significant effect on both attendance and learning. However, the programme has now ended, which makes schools question the sustainability of such programmes, including EQUIP-T. **All case schools perceive this question of pupils being hungry as central for improved attendance and pupil learning**. EQUIP-T may need to consider how to incorporate this in its design, in particular now when WFP is no longer present (at least in the nine schools visited).

8.4.3   Challenges to teaching

Infrastructural challenges mean that teachers face difficulties implementing what they learn in INSET. Respondents (including teachers, head teachers and SCs) feel that INSET does not fully consider the reality of the environment in which teaching and learning takes place. Though teachers believe they pick up skills around teaching, lack of classrooms and other resources pose significant challenges for them to be able to implement new skills effectively. Some schools perceive INSET to target urban schools more than rural ones, and they advocate for EQUIP-T to consider supporting rural schools with infrastructural development in order to have the desired effect. A head teacher of a rural/urban school said:

> *"After having made the training to the teacher they should change and focus on the situation in the schools, and their environments. For example, if you go to the village school you find that the schools are very old and the children are supposed to be taught in the same environment. Therefore, you find that the school [does not have] enough resources. Hence, after giving such trainings they should extend to the next step to support the rural village schools"* (School 3, district A).

However, all schools visited in the qualitative sample were rural, and as such these are the perceptions of teachers in these schools. When referring to INSET as targeting urban school, teachers consider the content taught to be tailored for an urban environment (or what they perceive to be an urban environment) with smaller classes and better facilities.

The mixing of grades within one classroom makes it difficult for teachers to teach effectively. For example, in cases where pre-primary and standard 1 students study together, teachers feel the focus on teaching has to try to include all pupils, which slows down teaching, *"the next class, you are hearing they are making noise. It is because they have been mixed standard 1 students with pre-primary to the extent that it disrupts the teacher's way of teaching because the teaching of standard 1 differs from pre-primary. But because they have been mixed, you will sometimes have to [teach] them vowels only"* (School Committee, school 1, district B). Mixing classes not only means having to 'cater to two audiences' but also for the class size to increase, which teachers feel makes them unable to fully include all pupils as well as control the class effectively. In these cases, teachers are thus unable to actively use the pedagogical skills they have learnt, due to a challenging classroom mixture. This mixing of pupils intensifies during rainy season, forcing schools that have opted for using outside space as a classroom to combine classes during that time.

Regional education officers are acutely aware that standard 1 in particular is overloaded. In addition to weak infrastructure, education managers often referred to teacher shortages as a key problem. Districts can put in requests for more teachers to central government but perceive the allocation to not be fair. Managers linked teacher shortages to the loss of instructional time due to

marking, the need for teachers to teach subjects in which they have not specialised or are comfortable, and the compulsion to teach in situations where each standard does not have its own teacher. All of the district education officers interviewed were frustrated about their shortages, relating it to the 'hardship environment' of living and teaching in these rural areas, as this quotation shows:

"*The situation on the teachers' transfers is not good in our District, maybe because most of the schools are in remote areas. For instance just last year, over 65 teachers got transfers away from the District. On each day, teachers are coming to my office with transfer requests.*" (DEO F).

Teacher shortages have recently been exacerbated by the reported increase in enrolment due to the removal of fees. According to an education manager:

"*Free education has brought in a disaster because many students are now enrolled whereas the number of teachers has not changed; teachers are finding it hard because they cannot divide themselves.*" (WEC X, district A).

The huge influx of children will make it extremely difficult for standard 1 and 2 teachers to put into practice the techniques they are supposed to learn from EQUIP training.

Lastly, language of instruction appears to pose a challenge to teaching, as some pupils are unable to understand Kiswahili, *"when these pupils start pre-primary and primary studies they do not know other languages rather than Sukuma so the process of changing these children from Sukuma to Swahili language which is the teaching language I think is a challenge to teachers"* (Community Leader, school 2, district C). Interestingly, teachers and head teachers rarely report this as a challenge until actively probed. However at the level above, education managers across all three districts mentioned children who only speak vernacular or local mother tongue to be a group who struggle with learning (sometimes linked to pastoralist communities), and cause a difficulty for teachers. In the majority of schools visited by qualitative team, it was clear that many pupils feel more comfortable with languages other than Kiswahili. In several FGDs, a number of our sampled pupils barely spoke Kiswahili, with other pupils translating words for them. Likewise, in FGDs with parents, in particular mothers, language was an evident issue and one FGD had to take place almost solely in another language, with one mother translating for the rest. In this same school, one of the teachers said he had to learn the local language since previously he could not communicate with his pupils. Moreover, schools in district C, which appears to have a greater issue with language than the other two districts, have introduced a 'punishment system' to control the use of non-Kiswahili languages during school time. Pupil's pass around a 'triangle-necklace' [bango] every time they hear someone speak non-Kiswahili. At the end of the day, this works as a reporting chain where the pupil reports who gave them the necklace. As one head teacher explains: "*by doing so, it is easier to identify the whole chain of Sukuma speakers*" (school 3, district C). As such, pupils are beaten if they speak their mother tongue at school. In one school, pupils said they no longer pass it on to help each other out and avoid corporal punishment. Moreover, language seems to be an even more significant issue in pastoralist communities, as Kiswahili tends to be less prominent amongst these communities. Consequently, as pupils are not always able to speak Kiswahili apart from when in school, and many pupils (in particular in pastoralist communities) might at times miss out on schools, there is limited exposure to Kiswahili for a lot of children. This may thus affect the time it takes for these pupils to learn the language, and as such affect their learning in the lower standards.

8.5   School readiness programme

Respondents who have a school readiness centre in their area can explain how the programme works[27]. WECs, DEOs and REOs felt the main rationale for establishing the SRP was to provide a 'pre-school' option for children who live far away from formal pre-primary and primary schools. As DEO D explains:

> *"We know that in this pastoral community the villages are scattered, [so EQUIP] established the SRP. What they are doing is, if a school is 10 km from certain community, street or whatever, we establish a classroom. […] Then identify the teachers from that particular community, the community itself, selected a teacher, then those teachers attended some of the trainings then went back to teach those kids."*

In this sense, the first aim was to make pre-schooling more accessible, with the second aim to help children be 'ready' for primary school in terms of learning in Kiswahili, some basic skills, and comfort with the environment. As one DEO said, *"to take them away from Sukuma that vernacular to Kiswahili that one, but also those learn life skills like to greet, to express themselves and whatever, that was the target"* (DEO D). To meet this language and skills need, the selection and tailored training of a community teaching assistant (CTA) from the local area is considered important. During the 12 week programme, the CTA organises the class through story-telling and play.[28]

At the community level, there was some confusion about the SRP and how it relates to the 'MEMKWA' programme.[29] Though some head teachers did know about the SRP, many respondents described a programme for over-age children to learn basic literacy, and referred to MEMKWA (Mpango kwa Elimu ya Msingi kwa Watoto Waliokosa). As one community member said when asked if they had a school readiness programme:

> *"The one that we are planning to start is the one that children cannot read and write and they are just roaming in street. These do not know reading at all, the class is known as MEMKWA, they want those children who did not go to school at least to know reading and writing. […Interviewer: Is the SRP different from pre-primary.] Yes, totally different. Pre-primary are the young kids but for the MEMKWA class these are children but of higher age."* (School 2, district A)

Thus there is clearly confusion in the community about these programmes, and they are not as well aware of SRP when it sounds similar to the MEMKWA programme.

After completing the programme, children of the appropriate age have entered standard 1. However some of the pupils were still seen to be too young to enter primary school: *"Others were not yet the right age to join grade 1; these are at home now."* (WEC 2 WEC Y, district C). These children were either below the government-recommended age or their parents felt they were still too young for the walk to the nearest primary school. Of these, some have remained at home, and some continued to attend the SRP even though it was supposed to finish (discussed more below). At the school level, teachers in one school explained that SRP children may then spend one year in pre-primary if they are not considered to be performing well enough for primary.

At the end of the programme, children are more prepared to learn, know how to play together, are more confident and comfortable in Kiswahili. Examples were given such as "*in the SRP centres, the pupils are taught how to speak simple Swahili. […] They are just playing but*

---

[27] This section is mainly based on interviews on district and regional levels, as the schools visited did either not have SRPs, or did not have sufficient knowledge to discuss it in depth.

[28] One WEC told us the programme was only run over eight weeks, and he did not know why it was shorter than planned.

[29] MEMKWA, or Complementary Basic Education for Tanzania (COBET), is an alternative approach for the provision of basic education to out of school children and youth.

when they join grade 1, it becomes easier for them to even master Kiswahili." (REO D); and *"the SRP children are like town children, they are active and they have courage of even asking teacher questions."* (WEC Z, district C). In this sense the SRP has made it easier for standard 1 teachers to include these children, who otherwise would not have had any pre-schooling.

However, SRP children are not considered to be performing as well as children who went to formal pre-primary schools. As a WEC said, *"children who went through traditional pre-primary schools are better off when compared to children from SRP. The children from pre-primary schools were taught by professionally qualified teachers"* (WEC Z, district C). Thus the CTAs are perceived as lower ability, with respondents pointing out they only had five days' training as opposed to the two years' training a qualified pre-primary teacher would have. On the other hand, one WEC noted that *"the SRP teachers focus only with the SRP classes. Techniques the SRP teachers were given are much better for the younger children,"* (WEC X, district B) compared with pre-primary teachers who are often busy with the primary classes in the same school. Another reason for the difference was the content: *"SRP has its own syllabus which is quite different to pre-primary schools syllabus."* (REO D). The SRP syllabus does not cover the full content of the pre-primary syllabus, and stakeholders note that children cannot learn as much in a 12-week SRP compared with two years' pre-primary.

KIIs showed that there are a number of positive spin-offs from the SRP. Some communities have asked the CTAs to continue their classes throughout the year, rather than finishing after the initial 12 weeks. In some cases this is for children who are still too young to join primary, but in others it is because the nearest primary school is too far away. For these centres, the intention is for them to become 'satellite schools,' as one WEC explained:

> *"A fully trained teacher goes to the SRP school twice a week to give more support. One SRP class is already converted to grade 1. The grade 1 pupils in the SRP class are enrolled in a mother primary school, however all sessions for the class are run in the SRP school."* (WEC X, district B)

By extending the SRP to also include primary classes, primary is now more accessible for children. This will have implications for the government, in terms of finding teachers to go to these satellite schools, and financial needs. Another spin-off is where the CTAs have been asked to support pre-primary schools by passing on their knowledge to pre-primary teachers. According to WECs, EQUIP is encouraging these developments.

District and regional stakeholders say, that where communities have the SRP nearby, the response has been positive. This is evidenced by the high demand for places in SRP centres: *"the centres were quite full of pupils; some of whom left pre-primary classes to join the SRP classes."* (WEC Y, district C). Centres have asked EQUIP if they can expand the class sizes, but *"it was not agreed."* (DEO E). The community and parents are said to be pleased with the SRP because it is close by, so their children do not have to travel a long distance, and it means their children are now in education. Some see it as *"a shorter way for pupils to start grade one as compared to pre-primary classes which are of two years."* (WEC Z, district A). Parents give financial or in-kind support to the SRPs: porridge for the pupils, contributions 'for soap' for the CTAs, and allowing use of other buildings or even contributing towards construction of classrooms.

However, according to these stakeholders, not all communities see the SRP so positively. Where communities do have the SRP, WECs report some cases of negative perceptions:

> *"[In] one village, the village chairman is not happy about it. […] The chairman says, instead of the children go to the centre to play and sing, they better remain at home to do some chores. The chairman wants the children start studies straight away. The chairman advised, the SRP should teach the children how to write."* (WEC Z, district C)

To this extent, respondents perceive some community members to not see a value in the methods used in the SRP, and for them to feel it is a waste of children's time to be in a centre since they are not learning academic skills. **Although this example of resistance only came up in one interview, it is an indication of the need for sensitisation about the programme**. Similarly stakeholders reported some communities to be confused about whether schools would accept SRP children now that pre-primary is compulsory, so this concern may prevent parents from sending their children to SRP centres if communication is not clear. This was echoed in responses by head teachers in the case study schools, who referred to SRP as 'pre' pre-primary, with a clear confusion around whether SRP children would be able to go straight to standard 1 or have to attend pre-primary as well under the new law.

8.6   Teacher motivation and morale

The EQUIP-T programme will not only seek to improve teachers' subject knowledge and teaching behaviours but also to improve teacher motivation and morale in order to increase teacher attendance, time on task, level of effort and commitment to the job (Cambridge Education 2014a).

8.6.1   Absence and punctuality

Teacher absenteeism from school and classrooms remain a major problem. Yet, teachers in all schools say that attendance has improved, partly due to *hapa kazi tu* (as discussed above) and to EQUIP-T training emphasising the importance of attendance and punctuality. This links with the head teacher SLM training,[30] and teachers consider the head teacher to be significant in changes to attendance and punctuality. This is particularly prominent in schools where the head teacher has recently changed (in 7 out of the 9 schools), and as such may not be directly related to SLM training but rather to a change in leadership. As one head teacher notes:

"It [EQUIP-T SLM training] has also supported us on small things that we thought were not essential, but they are very essential in academic matters. Like punctuality, this is because we are also trained on how to attend to the school early. Teachers self-awareness at the school also helped us on how to behave as teachers and time management" (School 3, district 1).

To this extent, some head teachers have introduced incentive systems for teachers to attend, such as a financial allowance at the end of each month if one attends all their classes (school 1, district A) and/or free coffee or tea (school 1, district A; school 1 and 3 district B and school 1, district C). Some head teachers are also encouraging teachers to improve classroom/school attendance by changing the system to request permission to be absent. For example, the head teacher in school 1, district B has introduced a two-day minimum notice if a teacher needs a day off, so that he has the opportunity to assign another teacher to cover the classes for that day. A school committee member verified this:

"The level of attendance is good because we now have another head teacher that has a different approach, so they come early and on time, and when he is sick there are no phone calls like in the previous years. The day before yesterday he even told me that there is a new approach of asking for permission for teachers who are sick by filling a form, not like the previous time when they used to make a phone call and others were lying" (School 1, district A).

---

[30] Discussed in detail in Chapter 9 on component 2: SLM

Yet, whilst head teachers and teachers self-report attendance to be high, and to have increased, qualitative observations, and pupil- and parent FGDs contradict this. According to a mother, *"some teachers are staying in…town, hence they usually come late and miss their first period"* (School 1, district A). In particular with regards to punctuality, several of the schools had teachers arrive late during the days the research team visited, and on one occasion the research team ended up picking up a teacher on the way to school. If the team had not done so, she would have arrived at school only on time for the second period, in spite of being expected to teach the first period. **All schools perceive the key reasons for loss of instructional time as workload and teacher housing.** Teacher housing is mainly referred to as affecting punctuality, though at times teachers say they may miss a full day due to weather or transport issues, *"…I am staying far from school and that forces me to leave early…sometimes you may wake up early but it is raining, or you get [into] a car accident on the way and come late to school. That means, if you have first period you will be late for 30 or 5 minutes…"* (Teacher, school 3, district C). Additionally, as previously mentioned, teachers may go home for lunch and then not return to school. Primarily parents mention this, stating it as a key issue, whist teachers only brought up the issue of lunch breaks upon probing. Interestingly, both teachers, head teachers, SCs, community leaders and district and regional respondents consider absenteeism to be justified in a number of instances. For example, according to a WEC:

> *"…teachers cannot be present due to different reasons. For instance, family matters, or they went to the district with concerns on work related matters, but not because of absenteeism. For instance someone has official reasons, family matters, and sickness not just intentional absenteeism."* (WEC Y district B?).

Teacher houses near the school could provide a solution, as this, in addition to teachers attendance not being affected by transport and distance, would also allow head teachers and parents to monitor more closely whether teachers have a 'real' reason for being late or missing school.

Teachers report unrealistic timetables and marking as making the workload too big, affecting both instructional time and their motivation to teach. Several teachers, head teachers, SCs and community leaders discuss the scheduling of teachers as too tight, with classes either overlapping or being consistently back-to-back, not allowing sufficient time for lesson preparation and marking. Though teachers rarely refer to missing periods, pupils say teachers are often absent from class, and that they spend time in the office 'writing' or 'marking'. For example, children in school 3, district B discussed what their teachers do while in school and one child said: *"some do teach one period only, and some do not teach anything at all…they are in the office…they are writing"*.. This is consistent with teachers' account of 'what they do in school when not teaching', though teachers state this to occur only during their free periods. However, qualitative observations support the findings from pupil FGDs, where teachers were seen marking or doing other work instead of teaching a class. When researchers probed teachers further, teachers would bring up workload with regards to marking as a reason for why they may miss periods:

> *"I start teaching from 8 to the end of the period. That is, you get out from this class you get into another class…So, there is no time for you to get free period because sometimes you may have so many exercise books to mark and you have to do it while missing other periods"* (Teacher, school 2, district C).

Additionally, though teachers may be present in class, it does not mean they are necessarily teaching. The qualitative team observed teachers in several schools assigning pupils work and then spending the entire period marking exercises at their desk. In one school, the teacher selected a pupil to teach the alphabet song, and then left the classroom to sit outside and

mark. Teachers highlight this as a challenge that has become more and more prominent as enrolment has increased, and that they have to mark and prepare more than before.

8.6.2   Monitoring

All schools perceive the increase in inspections to improve attendance, but some see inspections as demoralising rather than useful at times. All schools refer to both WECs and inspectors visiting the schools more often. Though WECs[31] conduct a variety of activities ranging from consulting with the head teachers to teaching classes themselves, respondents refer to inspectors solely coming to check on lesson plans and teacher attendance. However, teachers find that the increase in inspections are based on unrealistic expectations and do not support them to improve. According to a teacher:

> *"Personally, I am very upset with leaders at ward level such as WEC [and] district and regional management. When they come here they are supposed to provide critiques for improvement and provide directives 'here you did not do well' but they come here as police officers who come to arrest a teacher. 'Give the book', you give them… If they come here they have to give us advice, they have to see the burden that we carry and at least to appreciate what we have done"* (School 2, district C).

It appears that inspections focus on this monitoring exercise of inspecting pupils' exercise books. Monitoring the quality of lesson plans and schemes of work is considered a key role for head teachers and WECs. One WEC states: *"I first look at the scheme of work for every teacher. Another thing I am looking at is lesson plan for every teacher"* (WEC Y, district B).Teachers mention having to prepare lesson plans to show inspectors, and for their attendance to be monitored through whether pupils' exercise books have been marked. Linking this with the perceived need to mark more, and the increase in class sizes, there is a potential link with inspections and teachers' increase in workload. As previously discussed, marking seems to have an adverse effect on instructional time, with teachers at times prioritising marking over teaching. There may be a need to consider the potential risk of using marked exercise books as a way of monitoring teacher attendance, as this may in fact reduce instructional time in order for teachers to be able to meet these criteria.

However, there is also a perception, in particular amongst community members, that teachers are being monitored through the president's slogan *hapa kazi tu*. The increase in inspections is thought to be linked to the government's policy around 'working hard', and communities perceive this slogan to be the primary reason why teachers attend school more frequently. Teachers themselves also reference *hapa kazi tu*, and it is perceived both as a stick and a carrot, where monitoring is expected to either result in termination or promotion. As one SC member explains:

> *"Teachers' attendance is about 100%...they have to attend to this extent because of the current president's way of working. He makes follow ups on everything, therefore, if one does not teach [one] is in a position to be terminated"* (School 1, district C).

Though the qualitative data does not explore this in depth, **the perception of a wider change in policy around work ethic appears to significantly influence how teachers approach their work**, and in particular attendance. If this is a temporary 'scare' and does not result in teachers losing their jobs if they do not attend, or do not receive additional benefits if they do, the effect of this perception may not be significant/sustained.

Additionally, parents use their children's exercise books as a way of monitoring whether teachers are doing their job. Parents in all FGDs say they regularly check their children's exercise books to see whether teachers are teaching. Though teachers state parents do not pay

---

[31] See further discussion in chapter 8

attention to what they learn, especially in regards to checking on the content, is appears that at least some parents actively check whether the child has written, or has completed work on that day. For example,

> *"…attendance is good and for us parents we will know their attendance through the studies of our children. It means [that] if a person is making a follow up when a child is back home and you find that they have written, and if not you ask the child 'didn't you write today?', 'a certain teacher was not available', 'where did he/she go?' Maybe the child will tell you and you will say okay. But their attendance is good"* (Mothers, school 1, district C).

In this sense, teachers are being monitored both through visits from inspectors and WECs, as well as from a community level where parents are able to check teachers' work. However, this again highlights marking as a key activity that teachers need to undertake to 'prove' that they have been teaching. The question is if each classroom activity needs to involve writing and/or marking, or whether teachers can focus on other teaching methods in order to not always have a large marking workload. EQUIP may need to take into account the way in which assessing teachers attendance through pupils' exercise books may adversely impact on instructional time, as teachers spend a lot of time marking. INSET training could consider teaching teachers alternative methods for pupil assessment, and alternative means for monitoring teacher attendance may be needed, in the absence of additional teachers.

8.6.3   Job satisfaction and valuation by others

Respondents perceive teachers to be doing their work, but to not be motivated. All schools report that teachers do teach, and that this has improved in the past years. Yet, it is widely acknowledged amongst respondents that teachers are not necessarily motivated to teach but simply 'do their job'. The main reason for not feeling motivated appears to be the poor work environment, mainly relating to teacher housing (discussed below). Yet, EQUIP-T training is mentioned as having a positive effect on teacher motivation, with several teachers feeling more confident and able to try new things. This is mirrored in community perceptions around which teachers are motivated, with parents in FGDs across all schools considering teachers in standard 1 and 2 to appear more motivated to teach and to generally work harder, *"it is only the standard one teacher who works so hard…and he is the one who has brought changes this year…I think the previous standard teacher was not creative compared to the one that we have right now"* (Father, school 3, district C).

According to KIIs on a district and regional level, the profile, and therefore the motivation, of standard 1 and 2 teachers is increasing. In the past, teachers did not want to teach the lower standards because they lacked the skills, with teacher training colleges having apparently reduced the curriculum content for this group. As a result, Standard 1 and 2 teachers were seen as the bottom of the ladder, as if being there were a punishment. Now, by receiving training through EQUIP-T, teachers are more motivated to teach the early grades. One REO perceived this motivation to come from the attached allowances as much as the improved confidence and skills:

> *"Teachers are motivated, if you take them to a workshop they get some extra income, when it comes to distributing the subjects they are free to take. And because they have been trained, there is no option to deny that class, because they have been trained to do so. So they are prepared psychologically to teach that class"* (REO E).

Whilst the early grades used to be taught by older teachers who were nearing retirement, it is now more popular for those young teachers, *"after inception of EQUIP, it was learned that there are now more opportunities teaching early classes. In the past, the youth could say no to teaching early grades."* (WEC X, district A).

However, there is a perceived difference in benefits between being a teacher in urban areas versus rural areas, affecting teachers' willingness and motivation to teach in a rural school.

Some teachers perceive urban areas to have more teachers and smaller classes, and regardless of whether this perception is in fact correct, this directly affects their motivation to stay in a rural area. Teachers feel that they are not supported but unfairly treated in a rural school, and that they are not given the opportunity to meet the expectations placed on them, *"…we feel that the district isolates us for not allocating more teachers while there are more teachers out there"* (Teacher, school 2, district C). Inspectors are cited as a key reason for this feeling of unjust treatment, where inspectors are not considering the rural context when they 'judge' teachers performance and thus for teachers to be in an impossible situation to do well and receive encouragement:

> *"…when they come to us they yell at us and insult us, 'what kind of teacher are you, which college did you attend? You are stealing from the government'. So you look at it and ask yourself: why should I continue to be here? And we are only four teachers…I better ask for transfer to go to X place where there are more teachers and at least to rest for a while and reduce the workload as well as having five periods per day. If you ask them 'why does X place have so many teachers?' they will tell you 'we are judging a person in relation to this'. You just tell them, so long as I am lazy let me go to X place and those teachers in X place who are the best teachers should come here and after a certain period come to inspect me…"* (Teacher, school 2, district C).

To this extent, teachers see transferring as the only option, since they **perceive the government as unwilling to send them any additional teachers.** This comparison between the situation for urban and rural schools is prominent across case schools (which themselves are all rural), and teachers further feel that teachers in an urban setting are better placed to earn a living, since they are in a situation where they can complement their teaching income with other activities more easily.[32] **The feeling that 'no one cares' about their situational reality but continues to judge them makes teachers feel de-motivated to stay in a rural setting, and is cited as a key reason for their transfer wishes.** District and regional managers are aware of the challenges and complaints of teachers, and are themselves frustrated that allocations are made by central government and leave them with too few teachers.

All respondents in the case study schools (apart from pupils) identified teacher housing as the most common problem affecting teacher motivation and morale, and general willingness to stay at a school. Teachers consider work environment as the main reason for why they dislike their current position, with teacher housing being the main component of this (on top of workload and large class sizes). **Though both communities and teachers identify lack of teacher houses as an issue, there is a significant difference in expected standards between these respondents.** Whilst parents and communities emphasise: *"it is not a must for those houses to be nice, but they should just be many"* (Fathers, school 1, district C), teachers feel houses to not meet their basic needs, especially emphasised by teachers who have previously had electricity in their houses when teaching or studying in urban areas.

> "…*the houses that we rent are local houses called 'tembe' [a house whose walls and roof made from mud and sticks]. I think you have moved around and seen what it looks like. It means during the rainy season you may find your clothes are wet, it's really challenging"* (Teacher, school 1, district C).

Because of this, many teachers choose to live in town even in cases where there are houses for teachers available near the school, which, as discussed above, affects their attendance and punctuality (and even instructional time in terms of extra-classes).

Additionally, lack of teacher houses is the main reason cited as for why teachers request to transfer. This relates both to the cost teachers face in renting houses in a community where there is not much competition (resulting in higher rental prices), and for teachers to consider the living conditions below their standards, *"…when a new teacher reports to this school and see that there*

---

[32] Discussed in more detail below in the section on salary.

are no houses for teachers…they think of the cost of renting houses out of school…they decide to request for transfers to urban areas or in other places…" (Community Leader, school 3, district B). Teachers refer to the need for additional allowances in order to be able to live in town and afford transport to combat the issue. There is an inherent consistency here, where teachers do see the benefit of living near the school (attendance, punctuality, additional class hours, lower transport costs) whilst also feeling that they are settling at a living standard they are not satisfied with. All stakeholders agree that teacher housing is a main reason why teachers do not stay longer than necessary, in particular those teachers who have previously experienced more urban conditions.

Salaries and promotions

Low salaries, and lack of opportunities to increase these salaries, as promotions are scarce, demotivate teachers, *"when you get into the class teaching and you have no money at home…you are just thinking 'maybe if they could promote me I could still have money at this moment?...It really demoralises us and makes things more difficult and affect our teaching"* (Teacher, school 3, district C). Teachers link this to the monitoring previously mentioned, or the feeling of being 'overlooked', and argue that they do not receive the same chances as other teachers (urban) for promotions. As baseline also showed, stakeholders agree that salaries for teachers are too low to cover expenses. This further relates to several Head Teachers mentioning school expenses to have to come out of teachers' own pockets, and for Head Teachers to struggle paying them back on time, meaning that teachers are often out-of-pocket for expenses such as teaching materials. Respondents, including parents, say that teachers frequently have to take on other IGAs to support themselves. This is particularly true in cases where teachers have a family, and as such may be more likely to affect male teachers since there are often expectations for them to provide for a family. For example,

> *"The salary I get is not sufficient to cover expenses, and when you look [at] the starting salary of the teacher [it] does not match with the current situation in life. Sometimes you find some teachers doing other work other than teaching just to get money to cover expenses, because the money they get from salary is not enough".* (Teacher, school 1, district C).

All parents and community leaders spoken to state teachers to often have IGAs on the side, and for this at times to affect instructional time. In particular, during the harvest season, teachers may be absent from school in order to tend to their 'other business'.

The perception is that it is easier for teachers in urban areas to supplement their teacher salary through additional IGAs. Respondents believe teachers in urban schools are able to earn money through teaching extra classes (as well as having the time to do so, as there are more teachers to cover 'compulsory' timetables). As one teacher explains:

> *"Teachers living in town have so many opportunities. They can even do something to generate income. But here in the village you stay the whole day in school from morning to the evening without doing anything to generate income…they have to introduce teaching allowances for the teachers living in the village as it was before, to motivate them"* (school 1, district C).

However, teachers cite EQUIP-T to have changed this slightly, as attending INSET provides teachers with an opportunity to supplement their salary through the EQUIP-T allowance. It thus appears that teachers do not view these allowances as covering expenses for the training, but rather as additional payment for the 'extra work' they are doing, in a way viewing attending INSET as another IGA. In this manner, all stakeholders (case study schools and KIIs on a district and regional level) view salary and allowances as key motivational factors influencing teachers' willingness to both teach effectively, and to stay in a rural school. In line with baseline findings, stakeholders also view salary as having to change for teachers to put more effort into their work.

Community relations

Having to rent houses within communities appears to cause conflict between parents and teachers, at times relating to local norms. When teacher houses are not available, and schools are located far from a town, teachers have to rent houses within the local community of the school. Teachers and head teachers say this decreases communities respect for teachers, *"…the community that is living with them does not respect them. Taking an example, someone that has been educated and has a salary, yet they go to live in someone else's house which makes them not respected…"* (Head teacher, school 1, district A). Teachers renting houses, or renting rooms, thus appears to affect the power balance between teachers and the community, in a context where teachers seem to otherwise enjoy a level of 'authority status' as 'educated persons'. At the same time, education managers mostly emphasised how important it is to live with and be close to the community in order to understand them, and to show that teachers respect the community, *"what happens to the villagers, when they hear that you're not ready to live in their houses, what would they feel, they would say, this man feels to be nice more than us, can he/ she be there to our problems?"* (REO D). However one WEC felt very strongly that teachers should not spend too long living in one community, because it causes complacency and a perception that teachers have benefited from the community.

In a similar manner, respondents refer to teachers often arguing with parents whom they are renting from, in particular in cases where parents perceive a teacher to not have taught their child well. The perception is that this in turn affects a teacher's willingness to teach that particular pupil.

Additionally, both parents and teachers refer to teachers as mistreated in the communities, and in particular by their landlords, with landlords hiking up rent and monitoring teachers' behaviour. This seems to particularly affect female teachers, as many female teachers may come from urban areas and at times behave contradictory to rural, more conservative, norms. In one school in district C, the community leader referenced 'promiscuous behaviour' by female teachers as having caused conflicts between the school and the community. In another school in district A, the head teacher spoke of female teachers facing pressure to marry men from the community both to not remain unmarried (which is considered inappropriate) and to be able to afford paying rent in the community. Though the qualitative findings do not explore this in depth, **it is important to consider the varying challenges faced by male and female teachers with regards to living and working in the communities, and how this may affect motivation and morale to differing degrees.**

There appear to be conflict between communities and teachers with regards to corporal punishment[33]. Consistent with findings at baseline, corporal punishment appears to be the main reason for conflict between teachers and parents. Teachers feel they lack agency in disciplining children, and though parents are likely to use corporal punishment at home, they do not feel the teacher has the right to punish their child. Though parents in FGDs acknowledge that children may not be telling them the whole story, they still cite this as a main reason for grievance. This, often infected, relationship between schools and communities affects teachers' motivation and morale, as parents discourage them in their work or treat teachers harshly.

> *"In previous years, teachers were not confident in doing their work because of the bad relationship they had with parents. So you find a teacher saying [they] will not punish any child and just wait at the end of the month to get his/her salary"* (Mothers, school 2, district B).

However, the majority of schools perceive the relationship to have improved and cite new head teachers as the main reason for this, and for making an effort to improve community relations.[34] Additionally, **parents' attitudes towards education directly affects teacher's motivation to teach in a school.** Teachers in all schools state that parent unwillingness to push children to focus

---

[33] See further discussion in chapter 11 on Component 4
[34] Discussed in more detail in chapter 11 on Component 4

on their homework or be attentive in class makes it harder to teach and thus teachers can see little point in putting effort into teaching. In one school in district B, one teacher said he had overheard parents telling their children to fail so that they would not continue to another level. Similarly, a mother in school 1, district C stated that:

"A tribe like this of Sukuma is known [not to] like education…so they just do it to fulfil their responsibilities. Parents don't want their children to go to school. They are telling them to answer wrongly in the examinations so that they can [seem like] they don't know anything and don't move forward…We are asking [ourselves even] to understand the meaning of education. Because if we continue blaming the teachers while we are the ones planting bad seeds [so] a child [wont] listen to the teachers that means when he/she comes to class he/she will not listen to the teacher…" (school 1, district C).

To this extent, teachers feel unmotivated to teach as they feel pupils are not interested in learning, and even if they are, that the majority of pupils (in particular in pastoralist settings) will not be allowed to make use of their education. Yet, all schools cite this to have improved, and for parents to be more encouraging of education, which teachers in turn say motivates them in their task.

9 Qualitative evidence on school leadership and management

Component 2 of EQUIP-T focuses on the capacity of head teachers to lead schools effectively, ensuring that schools meet quality standards including those that are set for teacher performance. Effective school leadership and management is assumed to lead to increased teacher performance, which in turn links to improved education quality. To this extent one of EQUIP-T's key five outputs is to enhance SLM skills through targeted training for head teachers. The below discussion explores the evidence gathered around the role of head teachers as leaders, and their capacity to effectively manage schools across the nine cases included in the qualitative sample.

The high turnover of head teachers (discussed below) affected the depth of data collected at midline by the qualitative team. As head teachers in the majority of schools were new to the post, and had not received SLM as planned, the qualitative team could not fully explore the effect of EQUIP-T INSET SLM in the case study schools. As such, readers need to keep this in mind when considering the findings presented below, as these are in many cases 'baseline' data for these schools. Assuming that head teachers remain the same, end line will thus be in a better position to consider potential effects of EQUIP-T INSET on SLM.

9.1 SLM Training and head teacher turnover

Teachers, SCs, CLs and parents feel that the SLM training has made head teachers adopt a more inclusive and cooperative management style. Those head teachers who have received SLM INSET, and respondents who saw changes in the previous head teacher following SLM, consider relationship management as the main change, with head teachers adopting a more inclusive management style. This includes asking teachers for advice on how to solve problems, meet with teachers regularly, sharing budgets and plans with the wider community, attending community meetings and speaking directly to parents whose children do not attend school.

However, head teacher turnover in the case schools has been high. In seven out of the nine case schools visited the head teacher had been in the role for less than a year, with three head teachers promoted within the past two months. One of the new head teachers had previously been the assistant head teacher in the school, two had previously been a teacher in the school and the other four were completely new to the school. Some respondents cite previous head teachers leaving due to the difficult environment, and not wishing to live in the area. This is in line with reasons given for teacher turnover in Volume II Chapter 8.

> *"[I: why did the previous head teacher leave?] Due to difficult environment, so he asked permission to leave and he went near his home environment [to] be a normal teacher"* (Head teacher, school 2, district C).

However, the qualitative findings also highlight questions around whether head teachers are deliberately replaced or shifted for the purpose of improving SLM in lower performing schools. Several head teachers state having been shifted from 'better performing schools' to take up their current positions. As one EQUIP-T RTL said: *"we do reshuffling either to cascade good performance or to [provide a] more conducive environment to perform better"* (RTL D). It thus appears that head teachers are either strategically transferred, or choose to leave due to difficult working environments. The fact that head teachers choose to shift relates to the similar discussion in Volume II Chapter 8 on teachers: a lack of motivation and morale to teach in a rural environment with few facilities, which may be outside the scope of EQUIP-T to change. However, that head teachers are shifted for performance enhancement needs to be carefully considered when assessing the effect of the programme.

The high turnover of head teachers explains why some respondents mention that SLM INSET has occurred, in spite of the head teachers themselves not having received training. Instead, it has

usually been the previous head teacher attending the training. However, head teachers mention having received training on the school development plan (SDP) within the month before the research team came to visit (April/May). It thus appears that the previous head teachers have received the majority of SLM training, whilst the new head teachers have received at least one module on the SDP. This then means that **few current head teachers have received the amount of INSET envisioned on school leadership or management.**

All schools identify the importance of a head teacher with strong leadership and management skills in order for a school to run effectively. **Case schools also consider the role of head teacher as essential for other components of EQUIP-T to function**, citing the central role of the head teacher in managing relations between teachers, the school, and the wider community. As such, respondents (in particular teachers and SCs) consider it important for head teachers to receive SLM in order to aid school development, with one teacher stating:

> *"…since management is something tough, one may not be aware of her/his responsibilities. Therefore, if the training will be given, the head teacher will learn how to live with other teachers, what to do, how to manage things and how to make sure that we make excellence in academics"* (School 1, district A).

This does not necessarily mean that schools have seen the effect of EQUIP-T, as the new head teachers have not received SLM, but rather see the EQUIP-T training as desirable *as long as the current head teacher is the one to receive it.*

As such, all respondents (including WECs) feel the head teacher turnover to negatively affect the development of the school. If EQUIP-T trains head teachers who then transfer, skills learnt are lost, similar to how respondents feel about teacher INSET training. Whilst respondents thus perceive the head teacher to be at the centre for change, they consider this change less likely to occur, or to be less pronounced, due to the lack of sustainability. As discussed in one school committee FGD: *"the changes have not occurred a lot because we have a problem of changing head teachers every now and then. You can find a head teacher is being changed after every six months"* (School 3, district A).

In some cases, such as when the assistant head teacher took over from the previous head teacher, the skills appear to have transferred to some extent, with the new head teacher having learnt how to manage from the previous head teacher (who received INSET SLM). However, these new head teachers state that although they have seen what is important in SLM, they have not learnt how to implement it. Thus, even though some skills may transfer internally between head teachers**, the high turnover of head teachers is likely to affect the effectiveness of SLM training.**

Still, all respondents perceive better school leadership as a key reason for positive changes in the school. The majority of respondents attribute improvements in school performance to changes made by the head teacher, such as meeting regularly with teachers and communicating better with parents and the wider community. Respondents also highlight SLM skills as reasons why new head teachers are 'better'. Schools thus perceive changes to have occurred following the introduction of a new head teacher. As a member of a school committee says:

> *"We had two head teachers. That is before we had this one head teacher who is now being replaced with the new one who has not even finished two months. When we compare them there must be a difference. The previous one was not good in building the relationship between him and the teachers, compared to this new one who has brought changes"* (School 1, district B).

However, positive change in SLM is not necessarily a sign of the effect of EQUIP-T, as head teacher are new. Consequently, the midline qualitative data has not been able to effectively asses the effect of EQUIP-T SLM INSET. Whilst schools are positive about SLM training, respondents

see it as lacking sustainability. Given that SLM INSET is provided away from school (except for the SC1 module), the programme may consider offering more than one training opportunity to enable more head teachers to attend, and reduce the challenge caused by high head teacher turnover.

9.2   Head teacher capacity

9.2.1   Head teacher roles and responsibilities

The qualitative baseline found head teachers' understanding and implementation of their role and responsibilities to be weak, and to only be carrying out basic functions to fulfil identified areas of responsibilities. For midline, awareness of roles and responsibilities vary significantly across schools, with some head teachers displaying clarity in components of SLM, whilst others are uncertain of what being a head teacher really entails.

For example, academic leadership was identified by five head teachers as part of their role, and in many cases to have learnt that through EQUIP-T (either through SLM training, through INSET teacher training, or through 'EQUIP-T materials'). As discussed in Chapter 8 in Volume II on inclusive instruction, changes appear to have occurred about the thought process around children's ability to learn, with head teachers and teachers stating that they are now more aware of circumstances affecting why children may 'appear' unable to learn and of the importance to include these children.

Yet, as mentioned above, many head teachers were new to their post and had not received SLM training as envisioned. As such, the findings below need to be read in light of this, at times reflecting a baseline situation rather than a midline evaluation.

School Development Plans

All head teachers were aware of a school development plan (SDP), though this was not always implemented. As discussed in the section on SLM INSET, it appears that all but one head teacher had received some training or information around SDPs. During interviews head teachers would bring out the folder they had recently received, containing information around how to effectively create and implement a SDP. The majority of schools refer to having previous SDPs, though for EQUIP-T to have taught them how to make these more effective. Head teachers who had received prior training say they had already learnt the importance of SDPs, and had created them, but following the recent EQUIP-T training they plan to revise their SDPs.

> "…previously we were putting many plans which would end up not being accomplished, while now we have been taught to prepare a plan which may be split into small parts, and how to involve people in the plan that we have made" (Head teacher, school 3, district A).

In several schools, teachers and community members also highlight the importance of SDPs.[35] **Respondents feel SDPs make the running of the school more transparent, and thus to help create trust between teachers and head teachers, as well as between the school and the wider community**. As one teacher puts it: "*the school development plan is very good because it is more transparency*" (School 1, district B). As discussed in Volume II Chapter 11 on component 4, the school committee (SC) in all schools has a role in preparing the SDP, with the head teacher taking on a committee role in all case schools visited. As such, schools perceive the SDP to be a collaborative effort, with schools in many cases presenting the SDPs to the community for 'approval'. Head teachers mentioned this process to make it easier to later cope

---

[35] SDPs are discussed further in Volume II chapter 11 on component 4.

with parents' concerns, as they can point to the SDP and say *"remember what we agreed upon"* (head teacher, school 1, district C).

Teacher performance management

All head teachers (as well as other respondents) report the managing of people, data and processes as the central responsibility of head teachers. Head teachers mainly implemented this through monitoring teacher and pupil attendance. Compared to baseline, head teachers appear more conscious of the importance of attendance and punctuality, emphasising *hapa kazi tu[36]* and an increase in monitoring as the reasons for why they have started to understand their need to manage this. Additionally, head teachers who have received SLM INSET state one of the takeaways to be a new understanding of the link between punctuality and a 'good school', with a 'good school' identified as one with high academic performance. Yet, several head teachers state managing people as a key challenge they are facing, and emphasise the need for further training on this:

> *"…actually management is very wide. You need to have many perspectives. There are very many dynamics in people's change and perceptions. I think those people in charge will have to tell us new techniques on how to manage problems because techniques change. You cannot tell us to rely on principles or rules, because it is very complicated to deal with human beings as you know"* (Head teacher, school 2, district B).

This appears to be of particular concern to female head teachers, with two out of three[37] female head teachers saying they find it difficult to manage teachers. In particular, female head teachers appear undermined in cases where there is a previous male head teacher within the school, or where there are other older female teachers within the school. Neither baseline nor midline explored this gender consideration in managing of people fully, but there are some indications that gender may play a role in how effectively head teachers are able to manage their fellow teachers.

Head teachers are identified as key figures in ensuring teacher attendance and motivation.[38] Respondents in all schools perceive the head teacher to influence teacher attendance, and for head teachers to focus more on attendance than they previously have. Head teachers feel this has been a challenge, due to teachers' motivation being low. Head teachers have introduced various incentives to improve attendance, such as creating weekly reports on attendance, and motivating teachers through financial incentives if they attend all their classes in a week (Head teacher, school 1, district A). Additionally, the majority of head teachers say they try to lead by example, and for teachers to feel more motivated to teach if they see the head teacher attending and coming to school early.

Head teachers in all schools perceive their ability to discipline teachers to have increased. The regular monitoring activities by the district level, helps head teachers manage teachers as there are more transparent consequences if they do not attend or perform. Head teachers say they prefer to first speak to teachers individually if there is a problem of attendance, before they report anything to a district level:

> *"There are procedures that we have been given…the teacher that does not attend classes…I first call him/her and if he/she repeats I have to write a warning letter, and if it*

---

[36] Online discussion translates this as 'Just work' or 'Here it's only work'. See https://kabatilakanga.wordpress.com/2015/11/20/hapa-kazi-tu/ and http://thisisafrica.me/no-christmas-tanzania-hapa-kazi-tu/

[37] The third female head teacher (school 2, district A) had been teaching in the school for 17 years, and said that although she had found issues in the beginning, it was easier now due to being well known and respected in the school and the wider community.

[38] See detailed discussion around teacher attendance and motivation in chapter 8 on component 1.

> *continues I report and send them to our leader who is the WEC"* (Head teacher, school 3, district A).

Head teachers in all schools thus seem to appreciate the clear processes around teacher discipline, and feel it helps them enforce their authority on this matter. The process for managing poor performance of teachers is discussed further in Volume II Chapter 10 on district planning and management.

Teacher performance management varies significantly across schools, with head teachers not always fulfilling their responsibilities as academic supervisors. This appears linked to each head teachers own leadership and management style, as qualitative data does not show greater teacher performance management in schools where head teachers have received SLM INSET. Still, head teachers in all schools show elements of teacher performance management, and rather vary in the extent to which they implement methods. Consistent with baseline findings, head teachers consider checking lesson plans as the most important means of assessing teachers. Some head teachers also stated meeting frequently with teachers and to either have the academic teacher or themselves check teachers' lesson plans. In three cases, head teachers and other respondents also said that the head teacher at times observes lessons, to aid teachers in their work. For example, the head teacher in school 3, district C states:

> *"…as a head teacher I make follow-up for all teachers [on how] to use teaching aids and prepare lesson plans…I sit at the back of the class, and if there is a mistake I interrupt him or her by using English language and explain how it is supposed to be…we [correct them] in class and after that we have meetings twice a week…to discuss challenges in teaching and pupils' problems. Also, [if] in a certain subject [the] pupils' performance is low then I ask the subject teacher to tell us the reason why in his or her subject pupils' performance is poor."*

Teachers in one school reference how the current head teacher is better at managing them than the previous head teacher was, as he takes the time to explain what they are doing wrong rather than simply pointing out that they are doing things wrong. As one teacher states:

> *"the former head teacher has a power…we were teaching using experience and if you are doing wrong he will never show you the way, instead he will tell you 'you can go I will teach myself'…the current head teacher takes his time to coach 'do this, do this'…"* (School 2, district B).

However, most head teachers themselves find it difficult to instruct teachers on how to teach more effectively. **Head teachers consider it challenging both to know *how* to supervise teachers on the new curriculum, and to have the authority to do so in cases where teachers know more about it than they do because of having attended INSET.** For example, the head teacher in school 2, district B argues that as he has not attended EQUIP-T INSET on the 3Rs,[39] he is unable to supervise teachers effectively on how to use the new curriculum. Whilst he is able to coach teachers in general terms, as echoed in the quote above, he faces difficulties knowing whether teachers are performing in accordance with what they have learnt in training. In addition to challenges of not knowing the new curriculum, head teachers' own teaching responsibilities and other administrative tasks appear to limit their time available to actively supervise teachers. **To this extent, the majority of head teachers acknowledged improving instruction as part of achieving academic success, albeit all did not necessarily efficiently implement this.**

Community relations

The relationship with teachers and the community was widely highlighted as a responsibility of head teachers, by both themselves, teachers, and the wider community (parents, school

---

[39] 'The 3Rs' was used by respondents as a catch-all term referring to EQUIP-T modules on literacy and numeracy, and the new curriculum training.

committees, and community leaders).[40] In this manner, respondents perceive the head teacher as key in ensuring good relations between the school and the wider community, and consider this as a positive outcome of EQUIP-T training:

> *"[I: have EQUIP-T activities affected their work in any way?] Yes, they have [a] positive effect because there have been immediate changes after this programme started. The head teacher has been close to the parents, caretakers and the community in general"* (Fathers, school 1, district C).

However, though communities perceive these changes as due to EQUIP-T activities (since they occurred after the programme started), this cannot be asserted since head teachers are new. Regardless, **head teachers who actively involve themselves in the community are perceived as better leaders, and as mentioned above, 'openness' and 'cooperation' in management are strongly emphasised as key traits of a good head teacher.** The qualitative evidence around the roles and responsibilities of a good head teacher thus echo the assumption embedded in EQUIP-T's theory of change around community involvement and engagement and how it relates to effective SLM.

9.3 Factors affecting effective SLM

9.3.1 Head teacher motivation and morale

Whilst other stakeholders perceive head teachers to be motivated, because they are 'doing their job', head teachers themselves feel unmotivated and that they lack morale. Though the majority of head teachers feel committed to their job, and feel a sense of work ethic, they do not particularly like their job. As previously stated, the majority of head teachers have not applied for the role, but rather been assigned it, and three of the nine head teachers spoken to would prefer to be a normal teacher, as the workload is too large. Apart from the additional challenge of workload, head teachers perceive the same factors as teachers, of teacher houses and salary, to affect their motivation and morale.[41] The EQUIP-T component on SLM does not appear to take these issues of motivation and morale into account. The fact that head teachers often do not want to be head teachers, may affect their attitudes towards the role and in particular the workload associated with it. Whilst SLM training, or 'better' head teachers, may affect the effectiveness of the head teacher role, it does not necessarily aid in motivating head teachers. However, as the majority of head teachers are new, the qualitative evidence is not able to explore the effect of SLM fully.

Head teachers complain about workload, and face difficulties juggling teaching with administrative tasks and management. Head teachers in all schools state administrative tasks to interfere with their teaching attendance, and perceive this part of their workload to have increased in recent years. FGDs with teachers, parents and school committees further support this notion that it is difficult for head teachers to find the time to teach. This increase in workload relates mainly to an increase in monitoring, and the responsibility of head teachers to complete forms and report/send information to the district level. Head teachers perceive this to not only affect their attendance in class, but also their ability to go 'above and beyond' and follow up on wider responsibilities. As one head teacher states:

> *"As head teacher, my responsibilities are so difficult because there are many challenges I face. I need to teach, to work on administrative matters and at the same time do other school activities. This [makes it] difficult for me to make other school follow ups because [I] have so many responsibilities to do at once"* (School 1, district C).

---

[40] See detailed discussion around the involvement of community in school matters in Volume II chapter 11 on component 4.

[41] See detailed discussion around teacher motivation and morale in Volume II chapter 8 on component 1.

Similarly to teachers marking exercise books[42], **head teachers thus face a trade-off between tasks they ought to do, and struggle to find a balance between their day-to-day responsibilities, and their responsibilities relating to wider monitoring and support.** As such, though head teachers generally express an appreciation of district involvement and support visits by WECs, inspectors and other actors, they find the associated 'bureaucracy' to be too time-consuming and indirectly have an adverse effect on their ability to fulfil their role as head teacher.

Moreover, head teachers state they are expected to attend more meetings at ward level than previous years, and for this to affect both their attendance in school as well as their own motivation and morale. Additionally, head teachers need to bring school reports to the district level. Meetings frequently clash with class times, or require head teachers to leave the school early in order to reach the meetings on time. Moreover, transport is a challenge, and is often infrequent and weather dependent. Lastly, head teachers at times say they need to use their own funds to pay for transport, and for this to demotivate them to fulfil their responsibilities at hand. To this extent both head teachers and other respondents, including teachers and parents, believe that head teachers would benefit from transport assistance, either through allowances or by being given a motorbike, in a similar manner to how EQUIP-T has supported WECs.

Additionally, head teachers often have their families in urban areas (in particular if they have previously taught in urban schools), and thus commute, or live there on weekends – affecting attendance, travel time and morale. Head teachers in four of the schools have families in urban areas and commute to the school. This at times means that they fail to attend classes due to transport difficulties or weather. These head teachers further state that they would prefer to live with their families but are unwilling to bring their families to 'these rural schools' as the environment is not conducive, and they would prefer their children to attend school in an urban area. As such, this may affect head teacher turnover in these cases, as head teachers are unlikely to want to stay away from their families for too long. To this extent the issue of teacher houses also affect head teacher turnover, as it does for teacher turnover, and in turn the effectiveness of SLM training.

9.4   Sustainability and changes

The picture painted for the majority of schools at midline shows positive changes in head teachers' ability to manage schools effectively, with teachers and community members reporting head teachers to play a central role in school development. However, this may be due to the different leadership and management styles of new head teachers, rather than the effect of EQUIP-T. If the case is that head teachers transfer to other EQUIP-T schools, this may retain the value of SLM training. However, it is not possible to assess this through the qualitative part of the impact evaluation, as head teachers in the case study schools have changed. Still, seeing as schools emphasise the skills around people management (ability to manage relations between teachers as well as the school and the wider community) replacing head teachers may affect their ability to build ties with teachers and communities. This could limit the extent to which head teachers are able to implement and build on their new skills (since they will be starting from scratch in a new community).

Moreover, the lack of emphasis by EQUIP-T on head teacher motivation and morale may influence the effectiveness of SLM, as head teachers in many cases are not interested in additional responsibilities. Instead head teachers say they face a trade-off between their administrative/SLM responsibilities and their responsibilities as teachers. To this extent, the dual role of head teachers may need to be considered rather than seeing the 'head teacher' and the 'teacher' roles as separate units to 'improve'. There is a risk that head teachers will forgo one of their roles, as such potentially adversely affecting pupil learning regardless of which role is de-prioritised.

---

[42] See discussion in Volume II chapter 8 (on component 1).

10 Qualitative evidence on district planning and management

Strengthening the systems for district and regional education management is the objective of component 3 of EQUIP-T. Two levels of management are broadly targeted under this component: ward education coordinators (WECs), and district offices. Through the provision of motorbikes and grants, as well as training on school leadership (under component 2), the ability of WECs to effectively monitor and support schools is expected to improve. This component is also delivering two types of training for officers in local government authorities (LGAs, also known as a districts). The first is a course of modules aimed at building capacity on education management, and in 2015 the modules on strategic and annual planning were delivered, with further modules planned on budgets, monitoring and evaluation. The second is programme-specific training to support LGAs with their responsibilities in implementing EQUIP-T, such as managing financial decentralisation, WEC grants and PTP grants. EQUIP-T staff expect that districts would have better understanding of planning and that EQUIP-T activities would have greater prominence in the LGAs.

This chapter focuses on the perceptions of change relating to this component from the qualitative research. Interviews and focus groups were carried out with EQUIP-T staff, regional and district education officers (REOs, DEOs), WECs, and nine case study schools across three districts. The strength of evidence on changes relating to component 3 is limited for two reasons. First, the programme's focus on WECs and decentralised management has increased since expectations at BL, so the BL study does not always have sufficient detail to make comparisons at ML. However, the ML study presents an opportunity to build greater understanding of these actors. The second reason is that this component affects fewer, more senior people, and it was found that respondents at the school and community level were often too far removed from this component to comment meaningfully. This can often be a challenge with interventions affecting senior decision makers, and is exacerbated when these officers have incentives to respond in the way they think the interviewer wants to hear. This means there is less opportunity for reliable triangulation and depth of information than for other components. If Component 3 is felt to be a priority for the endline, the design of the qualitative research may benefit from substantial revisions.

10.1 Introduction to ward education coordinators (WECs)

By the time the midline was conducted, WECs should have directly benefited from EQUIP-T in three ways: attending training with head teachers (HTs) and assistant head teachers on school leadership and school quality standards, receiving motorbikes, and receiving monthly WEC grants to pay for fuel and attendance at meetings. In addition, they attended early grade reading and maths (which respondents refer to as '3Rs') and school readiness programme training under Component 1, and were trained as trainers for school committees and helped establish PTPs under Component 4.

The EQUIP-T programme is expected to help WECs become more effective at monitoring and supporting schools. This would be demonstrated by a greater understanding of their responsibilities, carrying out effective school visits, more frequent visits than at BL, and dealing with poor performance. In order to be effective, there is also an assumption that WECs are held accountable for fulfilling their roles. The rest of this section presents findings from the qualitative research relating to the inputs and expected outputs from the programme. The role of WECs is explored in more detail at ML than it was at BL since their role has taken increasing prominence in the programme than expected. This description and analysis thus serves as a reference point for making further comparisons at endline.

To preserve confidentiality, WECs have been labelled with the same district names as the schools but given a letter which does not link them to a relevant school. DEOs, REOs and EQUIP-T regional team leaders (RTLs) also have random codes to de-link them from schools and WECs.

10.2   WECs' responsibilities

The first responsibility of WECs, according to respondents from school and management levels, is to supervise academic matters in their wards. This includes responsibility for monitoring attendance of pupils and teachers, ensuring discipline, checking the school environment and monitoring and assuring the quality of education. Specific examples given for this are that WECs look at aspects of teaching such as lesson plans, schemes of work, self-assessments, use of teaching aids, and learning of the 3Rs. Within this WECs also coordinate examinations in the ward. This responsibility includes supporting head teachers and making sure they are also effectively managing the school. As one DEO said, *"To understand if the teacher teaches, provides enough exercises to the students, to make sure if the timetable is followed, class attendance has been done in the school. Or every report has been written and submitted to the HT to ensure that they have been done and report if there is any problem"* (DEO E).

There is a view that WECs should be helping to solve problems, not just reporting them. In this sense they should be supportive: *"if the head teacher fails to perform a certain activity, he assists him"* (School Committee, School 1, district A). All stakeholders should be able to approach the WEC to help resolve challenges, from the community to head teachers, teachers and pupils. Some teachers do feel able to do this: *"if the head teacher is unfair to us, [the WEC] is the first person we go for consultation"* (Teacher, School 1, district A). Likewise WECs feel it important to consult these different groups, to get a fair understanding of what is going on, and other respondents verified that this does happen: *"S/he normally asks pupils 'does your teacher teach you?'"* (Father, School 3, district A). Some community level members however describe the WEC's role as merely about collecting the information *"so that he can send the information to DED"* (Community Leader, School 1, district C) and it is then the District that will act on these reports. As one community leader said, the WEC's work is to *"identify the problems of the school and to take these problems to DEO so that these problems can be dealt."* (Community Leader – School 3, district A). Mostly, head teachers and teachers do report WECs as supporting them with challenges.

Quality assurers (previously called inspectors) appear to be seen as different to WECs in that they are expected go into more detail than WECs, and are supposed to provide greater feedback to support teachers to improve. As one WEC said *"My inspection is not in-depth as how school inspectors are doing. The inspectors are professionals therefore they are looking for much more detail"* (WEC Y, district A). Further, an EQUIP-T staff member explained,

> *"The quality assurer is […] the one who is mentoring teachers, to see how best teachers can teach. To see how best a learner can learn. After their meeting, they sit and discuss issues, bringing issues together, helping the teacher to find some actions or some alternatives, how to improve the class and so forth. But the WEC is the coordinator. Because he has 1 or 2 or 3 schools, and his/her role is to see how those schools are being supported by the HTs."* (RTL E)

The implication is that WECs carry out a more cursory monitoring practice rather than deeply assessing the quality of teaching and finding solutions with teachers. Yet one Regional Education Officer's description of WECs bears striking resemblance to the description of quality assurers above: *"[WECs] are supposed to inspect the work done, finding that maybe there is a shortcoming somewhere is his role to try to correct the students. We say the WEC has to act as the mentor to teachers in the ward"* (REO E). Whilst the difference between WECs and quality assurers was not explored extensively, there is a strong possibility that the roles are confused.

Moreover, WECs appear to play an important role with respect to teachers' welfare. Within the list of responsibilities, for example one WEC included *"…ensure good working environment for the teachers, ensure there is justice and that teachers are committed to their work and ensure teachers' welfare. And also encourage the teachers to apply for further studies and whenever possible, assist them in different ways"* (WEC X, district C). WECs help resolve challenges for teachers, and facilitate between teachers and the DEO's office.

More generally, WECs play a connecting role between schools and the district by taking information, directives and clarification on education policies to schools, and sending information and reports back to the district office. *"Their [responsibilities] include connecting schools and DEO's office, […] coordinating all education matters for primary and secondary schools within the ward, sending reports to District with different sort of information"* (WEC X, district C). In this sense WECs are the 'representatives' of the district office in each ward, and are seen as the 'boss' of the head teacher.

Additionally, WECs have a responsibility to connect schools to communities. Although head teachers have the primary responsibility for this, education managers see WECs as sharing this duty: *"Most especially the WEC and the HT, because WEC and the HTs are middle men, who are supposed to connect community and the schools."* (REO D). Here the activities focus around community sensitisation or dispute resolution, as one WEC described within the list of her responsibilities:

> *"…advise community on the importance of education and how to boost level of the education in the Ward; resolve education or academic conflicts within the Ward; responsible on enrolment and registration of children; sensitise community on education matters…"* (WEC Z, district C).

On the aspect of community sensitisation, WECs try to persuade parents of the value of education, and from this they encourage contributions from the community such as towards infrastructure.

> *"We had a public village meeting in the third village, whereby I sensitised the community to start construction of the class so that children from that village can start SRP classes."* (WEC Z, district C)

Although community members do not make the link between themselves and the WEC as one of the WEC's responsibilities, their own knowledge of WECs often does relate to this part of WECs' role: *"He was here last week encouraging people to contribute for school desks and pay a visit to school"* (Father, School 3, district A). WECs also have a responsibility for ensuring enrolment, registration and attendance of the children, which requires the support of parents. One WEC specifically said *"I am also responsible for making pupils' census/enrolment from the age of 0 to 12 years"* (WEC Z, district B), which requires follow up within the community.

10.2.1 Content of school visits

School visits are the main way in which WECs carry out their work, getting the information they need from the school. When they arrive at school, WECs usually begin by meeting the head teacher: "*he gets the first information from the HT and working on it then goes to the class and inspects individual teachers and students' work, so every teacher's work is being inspected by the WEC"* (HT, School 1, district C). WECs check attendance of pupils and teachers in the register, and a wide range of examples were given on how WECs then review the work.

Checking teachers' lesson plans and notes and schemes of work is one of the main ways WECs check the status of teaching. Many WECs are said to also check in pupils' exercise books to see if the lessons were delivered and assignments given as planned.

> *"To inspect lesson plans, lesson notes, children exercise book because teachers can prepare well but then don't give the assignment to the pupils; instead they remain in lesson plan, and lesson notes are not delivered to the pupils. So in order to overcome this he took pupils' exercise book and teachers' lesson notes and compares them."* (HT School 3, district C).

This suggests some WECs are diligent about understanding the progress of teaching.

When WECs visit classrooms to observe teaching, their focus is on use of teaching aids. Teachers confirm that WECs pass through the classrooms to check how teachers are teaching and children are learning, and WECs emphasise looking at teaching aids to check the quality of teaching: *"I look whether there are teaching aids, if there are not, I ask why there are no teaching aids"* (WEC Z, district C). WECs describe a specific and measurable checklist of the signs of quality teaching: "*For the teacher who performs well, first of all must have teaching aids, scheme of work, lesson plan and lesson notes"* (WEC X, district B), and within this teaching aids are most regularly looked for. The emphasis on teaching aids has come from the EQUIP-T trainings, and is a quick and visible behaviour to look for.

> *"[Interviewer: What impact did the 3Rs trainings have on the teachers?]*
>
> *As 3Rs teachers continue to get training, their motivation tends to be boosted up, their classes are full of teachings aids, they are capable of making the teaching aids, pupils are also instructed to make the teaching aids."* (WEC Z, district C)

WECs also seem very focused on monitoring the number of children knowing how to read and write when they visit schools. Many WECs talked about checking the numbers 'knowing the 3Rs,'[43] and whilst this was not explored in detail, the following passage demonstrates this focus.

> *"For instance when I go to the school, I look on how many pupils are there learning 3Rs. When I visit the next week, I also check whether the number of the 3R pupils is reduced, increased or still the same. We have to discuss with the teachers on what has happened, what have the teachers done to reduce the number of the pupils who do not know the 3R depending with the present timetable.*
>
> *[Interviewer: How do you check the number of children who have got the 3Rs?]*
>
> *In the first round I do assessment of the whole school. And make a list of all the children learning the 3R. When I go to the school next time, I check whether the list has come down. Sometimes, I also enter classes to check on the status because some teachers give wrong information."* (WEC Y, district C)

This emphasis on the 3Rs was also seen at more senior levels, as regional managers were able to report their numbers:

> *"[Interviewer: Which of the EQUIP-T activities do you think has had big impact in improving children learning?]*
>
> *Especially the 3Rs, has made big improvement. You know in my region, we have almost 40,000 pupils who could not understand how to read and write. Now it's almost 2,000."* (REO D)

This quotation also demonstrates that improvements in 'the 3Rs' are attributed to EQUIP-T.

Likewise WECs look for whether teachers are teaching in line with the syllabus. Some WECs said they check the teachers' work, in particular lesson plans and schemes of work, against the

---

[43] Whilst 3Rs refers to reading, writing and arithmetic, WECs only tend to focus on reading and writing.

syllabus. *"I look at the syllabus and compare it to what teacher has prepared and I know whether the teacher has done a good or bad job"* (WEC X, district A). WECs are aware of what teachers are supposed to be covering and their role is to monitor this.

Gathering information from different sources is important to many WECs, to verify what is happening in the schools. The first most obvious source is the head teacher: *"I have to rely on the HT, because the HT is with the teachers each day"* (WEC X, district B). WECs need to have a good relationship with head teachers to get reliable information, and in the case schools visited head teachers do feel their relationship with the WEC is satisfactory. However, WECs also identify a need for feedback from others, since head teachers can have an incentive to withhold problems.

A valued source of information is speaking to the pupils directly*: "I cannot by 100% rely on the HTs information…When I entered in class seven in the school, pupils complained to me that are not taught English subject"* (WEC X, district B). This practice is itself verified by the community, as one school committee member said *"[the WEC] asks the students if the teachers are attending their classes"* (School Committee School 1, district C). In addition to looking for feedback on teachers' performance, WECs use this method to check the level of learning of pupils for themselves: *"He calls the children and start giving them questions, 'come and write for me this or come and do that', he is checking on the ability of pupils if they are taught and have understood or not"* (Mother School 1, district C). Whilst WECs also say they talk to the community about the schools, community members did not mention being actively consulted by the WEC, so this may not be common.

Occasionally WECs teach classes, if they have the time and the head teacher and teachers agree. *"Not so often I teach, in most cases I teach when doing school monitoring. After I am through with teachers' inspection, I request session to teach, I like to teach Swahili and history subjects"* (WEC X, district B). This was confirmed by school level respondents from the same wards, suggesting it does happen but perhaps not everywhere.

Of course there are exceptions to the picture of WECs given above. This is best verified from the school level, since WECs have an incentive to respond according to what they think the research team want to hear, and may also have overestimated their own delivery in the role. In one school in particular, the teachers had a much lower opinion of the work of the WEC, who appears to be following the minimal processes required:

> *"Because whenever s/he comes to school s/he just does his/her work asking for scheme of work and lesson plan and check them and then leave. We cannot sit together like this to discuss together 'we have these challenges,' 'we would like you to do this and this'. There is no such a thing he only checks scheme of work and lesson plan and leave… Mostly, he just reports as evidence that he visited the school so that he cannot be asked 'why didn't you visit this school' and anytime he arrives in the school he asks 'give me the log book.'"* (Teachers School 2, district C)

Clearly in these cases the relationship is poor between WECs and teachers, and teachers will not be getting the support that they require. More senior managers also feel that some WECs are not performing, but that to some extent this variation is inevitable.

In addition to school visits, WECs talk about holding monthly meetings with their head teachers. This is an opportunity for heads to meet together and share their current strategies and challenges. These meetings allow peers to support each other and learn from each other, as one WEC said, *"I requested other schools to learn from [name of] primary school on the things the school did to improve its performance"* (WEC Y, district C). By setting a forum for heads to talk about academic matters, WECs feel this has changed the performance of head teachers, such as by borrowing from each other's ideas. Some WECs talked about 'ward disciplinary committees' which were made up of head teachers and the WEC, or 'ward education committees' which also included academic teachers – it is not clear if these are the same thing, and if they are standard

practice. These meetings were only mentioned by some WECs and reinforced by a few school stakeholders, so it is not clear if all WECs are holding them.

WECs face some challenges in fulfilling their roles effectively. Infrastructure and facilities are the main challenges identified by WECs and DEOs, as in the baseline. For example, there is often a lack of accommodation near the ward centre, and now WECs also need a secure place for the motorbikes. WECs lack facilities that would make their work easier, such as offices and computers, and writing reports by hand is considered arduous.

10.3 EQUIP-T inputs: motorbikes and WEC grants

WECs have received motorbikes, with some reporting that these motorbikes have been available from as early as June 2015. A small proportion of WECs had not yet received motorbikes because they were working in newly created wards or because EQUIP-T did not originally include the municipal districts. At the school level, whilst teachers knew motorbikes came from EQUIP-T, community members did not usually know.

Whilst WECs have been receiving the WEC grant, the payments have not been regular. A number of WECs complained that the grants were not timely. Whilst WECs received the motorbikes in June 2015, they did not receive any funds for fuel until January 2016, and had been expected to use their own funds until that time. *"Most of the time we have been using our own money instead of that money. For instance the motorcycles were received in June last year, since July, August, September, October, November and December this money was not given to us. We have started to receive just for the months of January and February"* (WEC Y, district A). WECs feel that the delay was caused by the district rather than EQUIP-T, as one suggested further monitoring from EQUIP-T could improve this: *"Maybe EQUIP-T should follow up to the district, in order for the funds to be on time"* (WEC X, district B). Delays in transfers were also mentioned by WECs in relation to the PTP grants.

WECs are expected to submit a spending report and a budget for the grant, and the understanding is that they receive a fixed amount. This understanding is held by WECs (*"As per EQUIP-T instructions we were supposed to be given 620,000 in a period of three month, first month 200,000, second month 200,000 and third month 220,000"* (WEC Y, district A)) and DEOs (*"That they are supposed to get 200,000 per month"* (DEO E)). Meanwhile WECs complain that their needs are different and as a result this average amount may not be enough. *"For instance where I am staying just one litre of fuel costs 2,500 shillings"* (WEC Y, district A). In fact according to EQUIP-T programme staff, WECs should receive an amount depending on their needs, so this would vary from person to person.

10.3.1 Changes due to the motorbikes and grants

There is general agreement that WECs are visiting schools more frequently than they used to. From teachers up to DEOs, stakeholders say that WECs now visit schools more often – ranging from twice a week up to once or twice each month – and parents too have noticed the increase. As one teacher said, *"in the past years it was very rare to see the WEC coming to join the school and the community but after he has been given the motorcycle he always come to visit the teachers"* (Teachers, School 3, district A). There are however exceptions, with some schools receiving systematically more frequent or less frequent visits due to their location.

> *"It's true that some schools are visited more. [Interviewer: Why?] For instance [name of] primary school is very far. In the middle there is forest, in this year, I went there just once because there was too much water. For the schools nearby the ward, I go quite often. However, I normally teach in the school located at the ward headquarter."* (WEC X, district B)

As such more remote schools may be receiving less support from WECs because of the difficulties in access, and this is particularly a seasonal problem with visits reducing further when it rains.

Provision of motorcycles has allowed WECs to visit schools more frequently. Senior managers speak of how grateful they are for EQUIP-T providing the motorcycles, and even parents say the increase in visits is due to the transport: *"I can say [things] have changed because I never saw him visiting the school but at least now he visits the school once per week, twice per week or once per two weeks by using a motorcycle, and we are even aware that he is in the school"* (Mother School 2, district A).

Many stakeholders feel that WECs are supervising schools more closely due to the frequent visits. To some extent this 'close supervision' may just mean more regular supervision, but there are examples where WECs look more deeply at issues because they can visit more often.

> *"When I visit the schools, one day is dedicated to grade 1, 2 and the administration, another day I visit remaining five grades with the administration."* (WEC Z, district C)

In this case the WEC picks different grades for alternate visits. Visiting often gives WECs opportunity to focus on specific challenges and be sure they have covered everything.

At the school level, teachers and community members perceive the more frequent visits and interaction to have improved the relationship between WECs and the school. In this sense, teachers feel that they know each other better and the WECs know what the current situation is in schools, *"so, it improves the relationship and he is close to teachers"* (Teacher School 1, district C). As one community leader said, the relationship had improved in the past two years *"because the previous WEC was always staying in his office but this current WEC is visiting his school every time and has a very short time to stay in the office"* (Community leader, School 1, district B). Conversely, some parents do not feel that the relationship has changed, but their relative distance from school activity may mean they are not aware of changes.

There are incidental ways in which the motorbikes have improved relationships between WECs and schools. First, the school no longer has to pay the WEC for visiting:

> *"…in the previous time if you had to call the WEC you have to ensure two or three things which were transport for coming and going back and you have to give something because they are working but now days they come at the school with the motorbike"* (HT School 2, district B)

As such the motorbikes and grants have removed a source of conflict between WECs and schools, which was identified at baseline. Second, motorbikes reinforce the status of WECs:

> *"In the past we used to have no motorbikes; when you get to the school, you're full of dust, tired, there was no respect at all; I couldn't look like an officer. However, when I go now, the motorbike has brought much respect."* (WEC Z, district C)

The provision of motorcycles thus can improve WECs' confidence and the respect from schools.

Data collection and reporting is perceived to be more prompt now that WECs have motorbikes. Senior managers value the more frequent visits because it means WECs can quickly collect data and report it up to districts and regions, and the information is felt to be reliable: *"nowadays EQUIP-T provided the WECs with motorbikes, the WECs are now having easiest way collecting the information"* (REO D). With motorbikes WECs can go to the district office more often for meetings and to submit data. For one WEC, the ease of transport made them feel more accountable to deliver:

> *"The WECs feel that, they don't have reason for reports to delay, because they are given the motorbikes and the fuel grant. Why should the report not be on time? Even if certain*

> *HTs are not bringing the reports to the WECs, the WECs find it easy to go to the schools to get the reports."* (WEC X, district C)

The motorbikes have removed a constraint, and transport is no longer an acceptable excuse.

There are some WECs who are not comfortable using the motorbikes. One of the WECs interviewed was struggling to use the motorbike as intended: *"Because of my age, I am scared to fall down. As I can remember, I fell two times from the motorbike"* (anonymised). This WEC was a woman close to retirement, and although this was not explored in depth in other interviews (other female WECs and older WECs expressed no concerns), it shows there will be at least some WECs who do not feel able to drive safely. This WEC in particular resolves the problem by asking teachers – or paying someone – to drive her. School committee members in this WEC's schools are frustrated that despite having a motorbike, she is not visiting as often as she could: *"…she was given the transport by EQUIP-T but instead they don't use those transport to visits their schools as how they are supposed to do"* (School committee, School 3, district B).

10.4   EQUIP-T inputs: Training for WECs

All WECs have attended trainings, but do not distinguish a difference between those targeted at school management and those under different components of EQUIP-T. WECs refer to various aspects of the school leadership and management (SLM) training, and are most quick to mention school development planning, which they had attended very recently.

> *"[Interviewer: What topics were you trained in?]*
>
> *We were trained on school management, preparations of school development plan, others I cannot remember although we attended trainings in many occasions. Also, we attended training on SRP and how to get the SRP teachers and the pupils."* (WEC X, district B)

Whilst WECs could not always give the specifics of training they received under the SLM component, they were often able to describe ways in which the EQUIP-T trainings more generally had changed their behaviour and ways of doing things, discussed further below. As the quotation above suggests, WECs associate their learning with all the trainings they have received. WECs, and more senior managers, feel it is important that they and head teachers attend all the trainings given at the school level so that they also know what they are supposed to monitor:

> *"Whenever the seminars are conducted, the WECs should also be included. For instance, when training 3Rs teachers alone say on counting, excluding the WECs from the training is not a good idea because the WECs are supposed to monitor the teachers. It is quite inappropriate for the WECs to monitor things they don't understand."* (WEC X, district C)

There were a small number of cases where WECs had not attended the SLM training. Whilst in one case this was because the WEC was new, in another case the WEC had been in post for four years yet was clear that he had not attended: *"I have not attended any training course on leadership from EQUIP-T. The trainings I went for through EQUIP-T with the HTs concern the modules"* (WEC Z, district B). Although this seems to be a one-off, it suggests a need for regular follow-up sessions to ensure that all WECs do receive the relevant training.

10.4.1   Perceived changes in WECs' capacity

EQUIP-T training has improved WECs' knowledge and ability to carry out their roles. Before EQUIP-T, WECs had received little or no tailored training for their role, and senior managers feel that previous experience as head teachers, and academic qualifications, are sufficient:

> *"[Interviewer: Do the WECs have the skills and knowledge to carry out their role?]*

> *I can say so. In [our region], the WECs are degree holders. But also I appoint those who were formally head teachers or heads of schools. Those experienced in teaching and administration."* (REO E)

Whilst WECs agreed that their previous experience prepared them, some WECs had not previously been head teachers, and many had only a certificate level qualification rather than diplomas or degrees, so these two conditions are not being enforced in appointing WECs. For these WECs, their confidence or perception of their own abilities does not seem any lower, but occasionally they will be disrespected by their better-qualified subordinates. One WEC, who had a certificate, said: *"For instance in one of my schools, I was given one HT who is a graduate. The HT felt degree was something very special. Because of the degree, the HT could not take any of my instructions"* (WEC Z, district C).

WECs themselves say that before EQUIP-T they had not received any clear instruction or guidance on meeting their responsibilities. When first appointed, they receive a letter *"with general information like you're going to supervise ward, academic, sending different information to District, such kind of information"* (WEC Y, district C). However, whilst this letter sets out responsibilities, *"how to do them has not been told."* (WEC X, district A). In this regard, WECs were having to work out how to fulfil their role based on trial and error, or guided by colleagues informally.

EQUIP-T training has given WECs a more structured idea of their responsibilities, and how to supervise and support the school. WECs talk about now 'knowing what to look for' when visiting schools, and one referred to this as a timetable:

> *"EQUIP-T does guide what to be done, where and when. Most of the things from EQUIP-T are done according to the timetable. EQUIP-T has really trained me to have good organisation, for instance having meeting schedule."* (WEC Y, district C)

The leadership training has helped WECs to structure their activities, but it has also made WECs aware of gaps in their own understanding of performing their responsibilities. This example from a WEC – who had been a WEC for nine years – demonstrates this:

> *"Yes, the training changed the way I am doing my works. Now I have wider understanding of different issues. Even when I go around the schools, I know what issues to look for as a result of the trainings. When I was appointed as WEC, I was not told about my responsibilities. However, through EQUIP-T I can now understand better how to perform my responsibilities."* (WEC X, district B).

This greater knowledge relates not just to the leadership training but also to attending INSET for teachers, so they are able to monitor teachers effectively.

Teachers noticed the change in WECs too, in terms of confidence, organisation and effectively solving problems.

> "*Another thing is that he is empowered to do his work and know his responsibilities. When he comes here he knows exactly what he is supposed to do compared to previous years where he was just given the position without knowing the roles of WEC. Today EQUIP-T has helped WECs to know their responsibilities and limitations are this and this and this. So, whenever WEC comes here he knows what to do."* (Teachers School 1, district C)

So whilst role of WECs has not itself changed, teachers feel the training has allowed WECs to be more effective in performing that role.

However, the change in government leadership is also perceived as having contributed to WECs' increase in commitment. WECs themselves see that there is more monitoring and supervision coming from central government and in turn the districts, and *"nowadays, there is*

much emphasis on meeting deadlines even from leaders" (WEC X, district C). Community members are quick to attribute behaviour change to the government.

> "*I think it is because of the new leadership, maybe that's why even this new [WEC] is coming to our school. Because the previous government leadership there were no follow ups made to the government workers like how it is being done now."* (Father School 1, district C)

This change, from October 2015, recurs across components in terms of increasing commitment, and as such makes it harder to attribute the effect of EQUIP-T alone from the qualitative research.

10.5   Managing WECs' performance

The EQUIP-T theory of change holds an assumption that WECs should be held accountable by the district in order to achieve high performance.

According to district and regional level officers, WECs' performance is judged based on the performance of the schools for which they are responsible. If a school performs badly, the WEC should be able to deal with it. Performance can be measured by a range of academic and administrative matters, for example examination results, teacher attendance, presence of school clubs and activities, or spending of capitation in line with guidance.

One way districts monitor the WECs is through a monthly meeting, in which WECs bring reports for all to discuss and then address challenges. Each of the district education officers feel this is a way for WECs to share their problems, but also for the district to see the WECs' reports and *"get time to do assessment of the WECs' performance"* (DEO F).

WECs do seem to feel more accountable to the district than at baseline, due to the resources from EQUIP-T. EQUIP-T has eased the challenges for WECs so there are fewer excuses for poor results. As one regional education officer said, WECs used to give excuses about lacking resources *"but now under EQUIP-T they don't have reasons to not be accountable for [poor school performance]"* (REO E). The idea that excuses are no longer acceptable was mentioned by both an REO and a WEC.

Despite this greater pressure, WECs are grateful to be receiving more attention due to EQUIP-T. One WEC, who had ten years' experience as a WEC, noticed this change:

> "*In the past the district didn't care about the WECs. It's like we were forgotten by the district. We had no working tools like chalks, stationaries, we were forgotten. Sometimes we used to buy them from our own money. Nowadays we are thankful on the coming of this organisation because it has given us the means of transportation. We are now working as a team."* (WEC Y, district A)

Meanwhile the district is holding WECs accountable more than in the past: there is a sense that punitive action is being taken more regularly, and this is affecting WECs' performance. Regional education officers say that demotions are happening, and becoming more regular: *"now they've seen we've started demoting they are taking seriously"* (REO E). REOs thus feel this is having an effect on WECs' efforts and improving their performance. As mentioned above, WECs themselves do feel that monitoring has increased. However, the source of this more zealous monitoring and accountability seems to be the new government and its focus on hard work.

10.6   Summary of WECs' changes

At the midline, stakeholders feel that EQUIP-T has allowed WECs to visit schools more often due to the motorbikes and grant, and that thanks to trainings WECs now have a better understanding of

what they are meant to do. WECs appear to focus on checking whether children know 'the 3Rs', look at the quality of teaching through lesson plans, schemes of work and marking, and look for teaching aids in the classroom. Teachers themselves note that WECs are now more confident.

Meanwhile, there are still some problems. Managers feel that some WECs are still not performing, but that it is natural to have some underperformance– there will always be some who do not perform, and are not suited to the job. This corresponds with the experience at the school level, where at least one school was dissatisfied with its WEC.

At the school level, WECs are seen to be more visible due to frequent visits, and more confident of what to do, so schools feel WECs are doing a better job. This increase in WECs' commitment is partly attributed to the new government, which is emphasising hard work and accountability.

10.7 Introduction to district planning and management

EQUIP-T's theory of change sets out that with the right information on schools, and capacity building on planning and budgeting, districts will become better at education management and evidence-based planning. At the time of the midline impact evaluation, the school information system (SIS) had not been rolled out, but district training had started and was due to continue.

10.8 Information and monitoring

One requirement for evidence-based planning and budgeting is that districts have the evidence and information needed. EQUIP-T is going to introduce a SIS, intended to improve accuracy and timeliness of information reporting. At the time of the midline research, this was just about to be rolled out, so at this point the districts had not benefited from the SIS but other aspects of the programme could have had an effect on districts' access to information from schools.

Schools provide a monthly report to the district which includes data on enrolment, gender, age, pupils' performance, teachers, infrastructure, school finances and challenges. In addition to monthly reports, there are the annual school statistics collected after each March: TSA, TSM and TSS (*Takwimu za Shule za Msingi/Awali/Sekondari* – Statistics on schools for pre-primary, primary and secondary respectively), and there can also be ad hoc information requests, according to needs from higher levels of government or emergencies such as diseases. WECs also produce a weekly report on their activities and ward. Head teachers up to regional officers mentioned this variety and number of reports. There is an upward process for information collation: head teachers must report to WECs, who collate and send to districts, then *"information may be compiled, and be sent to higher levels like region and PO RALG"* (WEC X, district C).

There are concerns at all levels of management that the information may not be reliable, and this is why verification is important. Education managers are frustrated that the data is unreliable: *"Sometimes, when you ask the number of students today, they give this. Then after two weeks, you request the same data and get a different number. When you try to check the reason, you don't get a convincing reason for the changes"* (REO E). WECs, district and regional officers see reasons why they may not be able to trust the information. Purposeful misreporting to benefit from the system is a major reason, best explained by an EQUIP-T regional team leader:

> *"if DEO calls HT, tells me the number children you have, the HT will like ask, is it for the school to donate money or receive money, you see, if it is for the school receiving money, big number will be provided, if it is for the school to contribute money, small number will be given."* (RTL D)

As in this example, inaccurate data can be because the school has some incentive to misreport, either due to financial consequences or to overstate performance (or hide bad performance). Other reasons are that heads give wrong data 'because of laziness, lack of seriousness,' not feeling it is

important to give truthful data. However there is also recognition that some of the data does change, making it hard to be accurate, particularly because of transfers of students or teachers.

The accuracy of information is felt to be improving, with strategies to verify the data being employed. Generally districts do not have the resources to visit schools and verify frequently, but WECs are supposed to verify the information, helped by the motorbikes to visit schools more frequently. WECs themselves feel that this is important to verify: *"If HT gave me wrong information, I will likewise submit wrong information to the District; all of us will be regarded as untrustworthy. What I am going to do is to make physical verification in order to be sure of the information"* (WEC Y, district C). Thus, as one mother said, WECs are *"afraid to be given false information"* (Mother, School 3, district A), and this relates to greater accountability, with WECs fearing disciplinary action if their information is found to be incorrect. Another strategy to improve accuracy is to request specific information, such as the names of teachers and pupils in order to verify the numbers: "*It is difficult to manipulate when the information is on the names"* (DEO F). Managers can then have more confidence in the data they have received.

Schools and WECs are not totally satisfied that the information they report is being used. Head teachers, school committee members and WECs either say they do not know what it is used for, or simply that "*it helps the District to know what is happening in the schools"* (WEC Z, district B) or to pass up to the next level. Respondents have rarely felt to see some change or action taken as a result of the data they submit. The main link that WECs and heads make is that pupil numbers are used for allocating the capitation grant, and whilst it tells the district where there are teacher, classroom or desk shortages, districts rarely have the resources to respond to these challenges. In this sense, WECs are frustrated at how frequently they have to send reports: *"when the same information is needed the next day, the district calls to inquire about the same kind of information. However, the information is there in the district, so they are not working on the information"* (WEC X, district B). Reports are seen to be repetition. This frequent reporting appears to be a change from baseline, as one head teacher said, *"I have seen there are now more assessments compared to last time, and the collection of information is very high in these two years"* (HT School 2, district B). Further research may be useful to determine what has driven the increase in reporting.

District officers are not able to visit schools frequently due to resource constraints. Whilst WECs have been provided with support under EQUIP-T, districts are supposed to fund monitoring from government budgets, which are unreliable.

> *"…access to schools by DEO's office, is somehow not good. Because access means you need to have fuel and the car. […] Previously we used to receive funds each month, we used to set certain amount for the fuel, certain amount for car service, but this time, I do remember the last funds came to us in December and this is now May."* (DEO D)

Districts – and in fact regional officers too – feel they now have to rely on WECs to visit schools and are frustrated they cannot verify for themselves. Schools and communities reiterated this need, as one community leader said*, "district leaders are not coming frequently and depend much on the information that is sent to them so it will be difficult to solve challenges"* (Community leader, School 1, district B). Communities and teachers expect that visits from districts would reduce bad performance and mean that districts understand what is going on, and the difficulties, in schools.

10.9   Planning and budgeting

Interviewees below the District office know very little about district processes. Community members and teachers do not know how the planning process takes place, or whether it has changed. Some WECs explained that the data they provide in reports goes towards the districts' plans, however the following conversation with a WEC, who had been a WEC for four years and a head for 14 before that, shows how detached some feel:

> *"Interviewer: Do you know how the District prepares its plans and budgets?*

> *WEC: Indeed, I don't know. I have not been involved to that.*
>
> *Interviewer: Do they ask you for input for their plan and budget?*
>
> *WEC: They do not.*
>
> *Interviewer: Do you think any of the reports you mentioned you're sending to them feed into the budget?*
>
> *WEC: Indeed there is not any feedback whether the information we're sending in is included in the budget."* (WEC Z, district B)

This is important in the sense that first, stakeholders feel detached from the process and it is clearly not transparent, and second, evidence on how the process does happen comes largely from interviews with only a small numbers of DEOs and REOs.

District officers know their annual plans and budgets should be made in a bottom-up process. Senior officers reiterated this, as one REO said: *"The schools do prepare the budgets, then taken by the wards, the wards send the projects to the District, the District compiles the projects and chose the projects they are able to implement"* (REO D). Stakeholders, including at the school level, explained that the final decisions are made by the District Councillors. Thus the budgets are said to be prepared by aggregating priorities and moving up the hierarchy, as is set out in the EQUIP-T training module on annual planning. Aspects of the module that were never mentioned by participants include the need for a situation analysis, the medium term expenditure framework, or the wider strategic plan.

However, the efficacy of the planning process is limited by the lack or delay of funds. District officers are weary with an annual planning process that sees them prepare a budget, be given a ceiling much lower than this, and still then receive less than their allocation.

> *"I don't know how next financial year's budget will look like, it is full of uncertainties. […] I don't have any assurance but in all the past years, you get less than what you budgeted and as a result the implementation becomes difficult. It may be that, you get a quarter of what you have budgeted."* (DEO F)

Districts are therefore in a position of uncertainty, expecting to receive less than they need but not knowing how much less. Further, funds can be late and unpredictable: "*So since January, we have never received any funds from the government"* (DEO E). Not knowing when funds will come, districts' plans for activities are disrupted. The delays and reductions are attributed to lower revenue collection than expected, or diversion to other priorities for the government. As one DEO said, referring to the EQUIP-T funds in particular, *"treasurer is like a pull, when the funds get in, they look for priority, like last year maybe I don't know where they took it for, maybe for election purposes"* (DEO D). Districts may be somewhat shielded from these challenges if they have their own revenue sources. According to REOs, urban LGAs may benefit from property taxes, and rural LGAs may benefit if they have natural resources such as mining. Further research would be needed to dig into whether districts feel this is a useful cushion for the education budget.

In reality it is hard to prioritise educational needs in this context. Despite the theory of the planning process, the LGAs' room for prioritisation is limited. First, priorities are imposed by higher levels of government: *"They [the districts] prioritise, the first thing which is very demanding in that particular year. For instance, our Minister here says that each student must have a desk, that's the first thing, the first project. […] The first priority demand comes from the policy of the government, but, the second choice comes from the community"* (REO D). The priority identified by the government may not reflect the main needs on the ground. Second, districts have to fulfill some recurring needs, such as administration costs and *"staffs' issues like health insurance, subsistence allowances, and even death"* (DEO F) without which *"when you fail to implement they bring you to trouble"* (DEO D). On top of that, emergencies can come up: *"Plans can change according to the*

schools' requirements. The school requirements can be very urgent. That's why the plan cannot be implemented as it is supposed" (WEC Y, district B). Each of these reasons, along with the unreliability of public finances, can squeeze the space for the district to carry out evidence-based planning and to prioritise based on the demands of its constituents.

10.9.1 Capitation grants

Schools have been receiving capitation grants monthly since December, with the funds sent directly to schools from the Treasury. Stakeholders note this is a recent change: "*[before the] election of our new president, we used to receive capitation once per quarter. But from December we receive capitation every month. And the money is sent direct to schools, not through district"* (DEO E). District officers, WECs and heads all say that funds from the government, or even 'from the president,' come monthly as opposed to once in every three to six months. Staff explain that the amount is based on the number of pupils, but they find it hard to explain exactly how much they are receiving per pupil. As such schools do not seem to know if they are receiving more now than before last December. Either way, education staff perceive that the president's commitment to free education has accounted for the change in payments to schools.

10.9.2 EQUIP-T's decentralised funds

In 2015 the EQUIP-T model changed such that funds are decentralised to LGAs. Funds are channelled through the government systems, shown in LGAs' development budgets, and LGAs have responsibility for implementing a number of EQUIP-T activities. This shift began with EQUIP-T allocations to LGAs for the 2015/16 financial year, which started in July 2015.

There is some acknowledgement at district level that the EQUIP-T planning and budgeting as a top-down process, which does not reinforce the bottom-up principles they are trained in. DEOs explain how planning for the EQUIP-T budget works:

> *"EQUIP-T is somehow a top down approach, […] because they have the statistics, they know that they have a certain number of schools, they know that they have certain number of HTs, they know that they have certain number of the WECs, so you might find that, they are planning for everything. […] What are we as LGAs doing, just implement what EQUIP-T has planned."* (DEO D)

EQUIP-T staff also referred to how EQUIP-T HQ does the budget, with LGAs providing data. As in the quotation above, this lack of autonomy annoys some managers, but others are less frustrated. At the more senior level, REOs are more positive: *"Then, we plan how to implement together, EQUIP-T members and the LGAs, even the regional secretariat we go together. We plan together, we agree together, so we sign it together that is our budget"* (REO D). The different view of REOs may reflect that they are more removed from the constraints of the budgeting process, and may be keen to stress the positives to the interviewers.

Centralised planning assumptions mean that districts find budgets do not always reflect reality. LGAs experience difficulties when the actual needs for implementation do not match the budget. For example, training participants are unhappy that the EQUIP-T allowances are lower than government gives, and a flat rate for travel costs is applied even though travel costs will vary. One district even had problems getting the facilitators with the allowances: *"They told me that, your schools are far from each other, when going to facilitate to the schools, bus fare may take up to Tsh50,000. They even told me that, the amount allocated for fare in the budget is very low. It's like they are not ready to facilitate [our district's] trainings"* (DEO F). Thus the unit costs set by EQUIP-T may make it hard for LGAs to run the activities. Another problem is that the model of implementation can change from the plan: *"For instance, we planned to conduct workshop for 3Rs, yet the budget could not include school inspectors. […So] we had to review the budget by doing re-*

allocations" (DEO F). As this case shows, over time the programme and districts themselves can decide the details of the activities need to change, and they then need to revisit the budgets.

There have been some difficulties with implementation since funds were first transferred in late 2015. Regional officers feel there have been *"some discrepancies which are not very important"* (REO D), passing off challenges as not concerning. However further conversations suggest that problems arise when looking at LGAs' spending as entered in the financial system 'Epicor.' As an EQUIP-T RTL said, *"Later on I came to understand that they were interchanging the codes in the allocation of the funds. Budget lines are overspent, under spent, not spent at all; others were spent but no activity"* (RTL D). Discussions with EQUIP-T staff revealed that the cause of the problem was unknown, and these misallocations could be human errors in the entry against codes (of the original budgets or the spending) or that activities cost more or less than originally planned. To some extent, the approvals process for spending should prevent activities going over-budget, *"it is my thinking that the Regional Administrative Secretary cannot approve the wrong budget"* (RTL D). However if this process is manual, there is still room for moving the funds.

LGAs have received funds much later than expected, and this is seen to be due to delays by the government. Districts understand that the central government receives funds for EQUIP-T, but then it does not release to districts immediately. As one DEO said,

> *"Another challenge which was found in this financial year, it is late release of funds. I do remember last year, I think it was in September, we were told that EQUIP-T has already released the funds to the Government, but from the September, the government released those funds in late December."* (DEO D)

This situation presents a challenge for the districts when they cannot implement as planned, and are then expected to condense activities into a shorter period. There is a risk that LGAs do not have the capacity to implement such a large load in a short space of time.

Districts feel that decentralisation has still increased government ownership of EQUIP-T. The responsibility now given to LGAs for managing funds, and regions' oversight role, means that EQUIP-T has greater prominence: *"…you know in the beginning of the programme, the district executive director and others were just hearing we are doing this, district treasurer and whatever. But now, they know that, they see it as very essential programme"* (DEO D). EQUIP-T is now known by officers outside of Education, as it is a source of funds for the LGA. The transfer of responsibility may also have increased programme sustainability, as one DEO said, *"…because the ownership of EQUIP-T has been given to us. Since they have changed the way of disbursing funds to us. And it means that they have given us the chance to prepare before the programme ends"* (DEO E). Government staff now have direct experience in organising the activities. DEOs recognise that continuation of activities will depend on the budget from the government.

10.9.3 EQUIP-T training for districts

District officers feel they have gained useful knowledge from EQUIP-T trainings. Only one of the three DEOs specifically mentioned skills on planning:

> *"The training capacitated me a lot. […] I used to do the budgeting based on the guidance from my boss. The boss instructed me how to go about it as I was not able to attend any training. […] Something I gained, which we used not do is that, when doing budget, you have to get inputs from the grass roots. […] Because the process captures many needs, you then have to prioritise the needs, on this; the training helped me quite a lot."* (DEO F)

This DEO could clearly explain what she felt to have learnt from the EQUIP-T training, and this had helped in performing the job. This perception was reinforced by an REO: "*According to them when they did the planning and budgeting they said it was easier than years ago. Years ago they were just copying from somewhere. But now they are saying, they knew at least what they were doing*"

(REO E). Another DEO said they had not attended any modules on planning and budgeting, but had been to trainings on *"the way to plan activities run or done in EQUIP-T programme"* (DEO E). If either the DEO had forgotten about other trainings, or had not attended them for some other reason (in this case they had recently been transferred but from another EQUIP-T district), the modules on planning may need to be repeated and reinforced.

Trainings organised under other components have also built DEOs' capacity to manage. As one DEO said about the trainings, *"when I go to schools, I always go there to assess what I was even learning there, I go there competent knowing that what I am going to assess"* (DEO D). In this sense, whilst components 1 and 2 may have been intended to strengthen teaching and school leadership, they are also important for more senior managers to know how to monitor education quality. This reiterates the point made about WECs earlier.

The turnover of district staff creates a risk for the effectiveness of EQUIP-T training. Of the three DEOs interviewed, two had been transferred from other districts in the past four months. Although their previous positions had been in other EQUIP-T districts, this may not be the case everywhere.

10.10 District management and relationships with teachers

Another area of districts' responsibility is managing the performance of teachers, schools and WECs. Although this has not been explicitly addressed in the EQUIP-T interventions so far, it is an assumption necessary for WECs, head teachers and teachers to be held to account. At the same time, positive reinforcement and incentives are important as well as negative penalties.

Districts have tools to incentivise good performance. In district B, schools receiving the lowest exam results are given a black flag. This was found to motivate teachers: *"after that [they] decided to reorganise themselves"* (School committee School 2, district B), and a WEC also describes how it prompted reactions and performance of a school improved by the next year. In district A, each ward gives a trophy to the lowest performing school, called 'ngao ya ujinga,'[44] which creates *"a sort of competition in teaching"* (Mother, School 3, district A). This same district also gives certificates and monetary prizes for schools and teachers who show good performance. It is not known if these examples are standard across LGAs or at the initiative of the district office.

Poor performance of an individual should be dealt with by progressive layers of the management chain: from teachers, to heads, to WECs, and then the district office. This route of escalation was described by school and district level interviewees, and the following passage from a WEC describes how this was used to deal with a teacher with poor attendance.

> *"The teacher was out of the school for one month. The head teacher warned the teacher, still the teacher could not change; the matter was then escalated to me. I had to call my ward education committee; still the teacher could not change. I reported the teacher to DEO's office; the teacher was given a warning letter. In the letter, the teacher was asked to explain why they had not been to work all that period. The teacher eventually changed their behaviour, the DEO allowed the teacher to continue working."* (WEC Z, district B)

As described here, the proper way is to work through the layers, even sometimes reaching the region if a problem cannot be resolved by lower levels.

When a teacher or head teacher is struggling with one of their responsibilities, WECs may first try to support them to resolve it. This might be where at teacher is not comfortable with a topic, or a head is having problems with bad behaviour from a teacher. Resolution might come from the WEC's own advice or arranging some other support, as one WEC said:

---
[44] Google translates this as 'shield of ignorance.'

> *"For instance I had one HT in one of my schools, the HT was not any leader before. So I decided to coach him very closely, every time I came to monitor how he does his job. Sometimes I request more experienced HTs to guide him."* (WEC X, district A)

In this sense WECs talk about 'coaching,' and heads value this support from WECs: *"I cooperate with the WEC in different things especially those which are difficult to solve by myself on the school level, he comes and helps me out."* (HT School 1, district C). Hence where performance is low because of capacity, superiors first try to advise and help the person, and that help is valued.

When poor performance persists, punitive actions can be taken by the DEO. WECs and ward education committees can only give warnings, but do have some leverage in advising districts on what action to take: *"…I report the teacher to DEO and request the DEO to transfer the teacher"* (WEC Y, district A). LGAs have the authority to take the most consequential action to address poor performance, and transfers and demotions appear to be most commonly used. WECs feel that transfers happen regularly (and as above, even request it), but more senior officers report that they should now only happen to *"cascade good performance"* and that *"our president does not allow transfers and he said this several times"* (RTL D). Instead, there is a focus on demotions: *"By November last year, I demoted some of the teachers, I mean HTs and WECs because of [absenteeism]"* (REO E). This suggests district and regional officers are using demotions more regularly now to deal with poor performance. Another penalty used is withholding salaries until performance improves. Redundancies seem to be very uncommon, with only one EQUIP-T RTL mentioning it and as one WEC said, *"in most cases, the HT won't be sacked from the employment"* (WEC Z, district B). Sacking seems to be limited to very extreme performance issues, with one WEC referring to sacking in the case of teachers sexually abusing pupils.

The visibility of demotions and penalties has increased accountability. First, staff are making sure to improve their own performance, but also this increases the importance of supervision for staff who have management responsibilities.

> *"Of course it is everywhere now, people are conscious to be either being demoted or taken off the office. In order to stabilise my chair, I should make you accountable because my performance is based on you. If you don't perform, I am not performing, so I cannot allow you not to perform, I will make sure you perform. So I will come day to day to see how you're performing, which challenges are you facing, can I help you?"* (RTL D)

Thus at each level there is now more incentive to improve the performance of those being supervised. Regional officers had seen this change, and WECs reinforced this with their need to verify the information they are given.

Relationships between different levels of the education system are felt to be improving. To a large extent, this is due to more regular contact, and examples came from the WECs attending training with districts, and the districts attending teacher training:

> *"In EQUIP-T trainings, we get a chance to meet with the district guys, in such meetings we get chance to discuss various issues in our wards."* (WEC Y, district B)

> *"also those frequent trainings which have been conducted for standard 1 and 2 teachers, HTs, INSET coordinators. […] so you can find an increased number of teachers whom I meet frequently."* (DEO D)

These trainings give an opportunity for managers to get to know their staff and their challenges, and for junior staff to then feel more comfortable to approach their superiors. Similarly, the regular meetings between districts and WECs, who bring reports from the schools, is seen to help improve the relationship: *"because [the WECs] report to us every month whatever going on in the schools, it has come to a point even us, we are now at least more close to the teachers than before"* (DEO D).

However, whilst teachers and head teachers noted the benefit of more frequent interaction with WECs, they did not mention any improvement in their relationship with the district.

10.11 Regional role in management

Regions have not so directly been the beneficiaries of EQUIP-T, and are not a critical step in the TOC. Nonetheless, EQUIP-T staff emphasised the role regions have in overseeing districts.

Regions may be involved in resolving problems when they are escalated up. Following the hierarchy discussed above, the regional office may eventually be called upon. As one REO said, *"the villagers […] telephoned me saying, we told DEO that the HT is not doing his work, we have now decided to telephone you as well, please come to our school and take your teacher"* (REO D). This only seems to happen when other routes have been exhausted, and the REO is then in a position to direct the DEO to take action.

In managing EQUIP-T, regions play a role in agreeing activities with the EQUIP-T regional team leaders, and communicating these directives to districts.

> *"The regional education office is in charge of us […] It is a regional academic officer who sends us report on what we are supposed to do together with EQUIP-T team at regional level. All of them they work together to tell us to prepare what is needed at that time. So they work hand by hand."* (DEO E)

Since regions are above districts in the government structure, their involvement is important.

Likewise, regions are supposed to be active in monitoring the districts' implementation. The regions collect monthly and quarterly reports from LGAs on their activities and spending. However regions feel that depending on the reports is not enough, and it is important to visit districts and schools to verify: *"After collecting [the reports], we need to analyse them. But sometimes you have to go there to sample, to have some areas you visit physically. So when you try to defend what you have received and you've analysed, you know them"* (REO E). Regions, as with districts, are frustrated that they do not have funds for visits. *"I don't have reliable budget which it could enable me to buy fuel, in order to make supervision, especially to the EQUIP-T implementation. If I need money for supervision, I have to ask the regional EQUIP-T coordinator to offer me fuel, to offer me allowances, in order to make supervision"* (REO F). For some REOs, this clearly limits the monitoring they are able to do. However another REO joins monitoring trips with the EQUIP-T regional team: "*And we are going together with EQUIP-T, when we were going for monitoring and evaluation we are together"* (REO D), and in this case sees this as strengthening the relationship between government and EQUIP-T. Regions' ability and enthusiasm for carrying out physical monitoring clearly relies on a positive relationship with EQUIP-T and some funds being accessible.

10.12 Summary of districts' changes

District education officers have participated in EQUIP-T trainings, and those who have attending training on planning feel it has increased their understanding of the planning process. However, districts' ability to use this process in practice is limited by systemic challenges. Districts continue to be allocated – and then receive – budgets far below what they feel they need, and after taking account of government priorities and office running costs, there is no room left over for prioritisation of schools' needs.

Districts need reliable data from schools if they are going to plan based on needs. The quality of data and reporting is felt to be improving, in part due to the additional verification being carried out by WECs, but also an increased sense of accountability. Districts and regions appear to be taking

punitive action – including demotions – more frequently than in the past, and this has made staff more committed.

The EQUIP-T decentralised funding mechanism has increased a sense of ownership of EQUIP-T in the districts, as it is a source of income for LGAs and places responsibility on the district. There have been challenges with implementation of EQUIP-T budgets, due to unrealistic budgets being set which do not reflect the local implementation model, and possibly due to capacity difficulties with using the financial system. The late release of funds to districts has particularly disrupted LGAs' implementation plans. Nonetheless, the fact that districts are experiencing running the activities themselves may make continuation of aspects of the programme more sustainable.

11 Qualitative evidence on community participation and accountability

Component 4 of the EQUIP-T programme seeks to empower communities 'to take an active role in improving school outcomes and services by equipping them with tools, resources, and the provision of appropriate, accessible information. The theory underpinning this component is that the greater engagement by the community in school operations and outcomes will provide much needed support and resources that will enhance the quality and relevance of education.

11.1 Communities' capacity-building: School Committees

The qualitative research has found that the schools visited consider school committees (SCs) the communicative link between schools and communities. Indeed, the majority of the respondents both from school and community level refer to SCs as *"the voice of the parents"* (head teacher, school 3, district C) since SCs are composed mainly of members of the community, whom are selected by community members who often are pupils' parents. The committee also includes members who enter it through *"their status"* (Community Leader, school 3, district A). These include, at school-level: head teacher, school bursar and teacher representatives (selected by teachers) and, at community-level, members from the village committee (selected by the village government). One member from the community takes on the role of chairperson, whilst the head teacher is usually the secretary of the committee. As seen at baseline, all schools visited have SCs established.

11.1.1 EQUIP-T training

The content of EQUIP-T training, according to the theory of change that underpins it, should address SC roles and responsibilities, financial management and Parents-Teachers-Partnership (PTP)/PTP grants[45]. **Whilst all SCs across districts have received some training, helping them achieve EQUIP goals, it is less clear when SC members are referring to EQUIP-T-provided training.** SCs appear to have received training from different actors, as one community leader mentions: *"this SC got INSET from government, EQUIP and millennium project so in general they know how to perform their responsibilities"* (school 1, district B). This makes it difficult to determine whether informants are referring to EQUIP-T training or to training provided by other actors. **There seems to be a lot of confusion around which programme provides what.** It is also unclear whether all SC members attended these training, as well as what content the training cover. This lack of clarity seems to be due to issues of definition, uneven implementation, and the existence of multiple actors providing training. Moreover, it is important to acknowledge that there may be an element of social desirability, in which respondents want to show what EQUIP-T has done in order to cooperate with and help the research team in their research. As such, respondents may at times assign components to EQUIP-T that other actors such as the government or the millennium project are in fact responsible for implementing. To this extent, **any analysis on SCs' awareness of their roles and responsibilities and capacity building needs to consider the various trainings SCs might have received.** As such one cannot necessarily attribute change to EQUIP-T.

However, **even when SCs assert to have received the training explicitly from EQUIP-T, they perceive it to not fulfil the full purpose of capacity building, but rather to be an informative lecture – with further training needed on how to be effective as a school committee**. Some SCs in FGDs distinguish between seminars and trainings, with seminars seen as brief meetings where SCs can discuss school issues and solutions, and trainings as more structured learning opportunities. SC members perceive seminars to provide knowledge around an issue, whilst

---

[45] Discussed further in section 11.2 on PTPs.

training teaches you how to implement it. In many cases, SCs appear to see EQUIP-T training more as the former, referring to it as seminars, and voice wanting further actual training. One school committee member states:

"*…as the member of the school committee [am I not] supposed to get training? [Training] would [inform me of] my responsibilities, [for example] monitoring teachers' [and students'] attendance. [I have been a] school committee member [for 2 years], but I don't know my responsibilities…That is why we are suggesting that there should be training that will give us the skills*" (school 2, district A).

Thus, though SCs may have received EQUIP-T training, they do not see this as adequate enough to be defined as 'training' according to their definition. SC members feel more targeted trainings necessary in order for them to understand their roles and responsibilities as well as increase their capacity to carry these out.

All stakeholders mention training or seminars as one-day-long single occasions. Whilst the majority of stakeholders among SCs and community leaders perceive one-day-seminars to be insufficient for effective capacity building, it does appear to create a space for SCs to come together and discuss school issues. **The qualitative study shows that either multiple training modalities are in place or respondents do not have a common knowledge on what to expect from EQUIP-T training and therefore they cannot articulate its modalities.** In one school, a member of the SC mentions a book provided by EQUIP-T as an effective way that has helped SC to understand roles and responsibilities. In another case, there seems to be a cascade training model, where the head teacher first receives training with the WEC, after which they "*come back and teach the SMC, so they perform their duties well*" (head teacher, school 3, district C). **Respondents in SC FGDs appear unclear around the specific learning components and main takeaways from training.**

11.1.2 School Committees: roles and responsibilities

SC members and other stakeholders perceive SCs to have an overall leadership role in the school management like "*a father in the family supervising all school development activities, solving school challenges*" (Teachers, school 1, district C). According to case study schools, SCs' main responsibilities are:

- Addressing school needs and solving problems within the school and between the school and the community (acting as a bridge between the community and the school involving parents/community in identifying school problems and finding solutions, and representing community/parents' needs through school meetings);
- Managing school budget, facilitating Income Generating Activities (IGAs), and supporting school development plans (SDPs)[46];
- Ensuring pupils' attendance and better academic performances through monitoring (educating parents on the importance of education); and
- Monitoring teachers' attendance and commitment.

In terms of addressing school needs and managing school budgets, respondents say that it is now mandatory that SCs approve all the proposals for how to use money for school improvements (i.e. maintenance of the desks, building teachers' toilets, purchase of school materials) before being implemented. Additionally, SCs do not only manage school budgets, but they also supervise activities that generate income by deciding, for example, which farm activities to carry out in the school's premises and making follow ups on these. Therefore, SCs sit in meetings together with teachers to evaluate different suggestions. During these meetings SCs represent parents' voice

---

[46] See section Section 11.3.2 for further discussion on SDPs as well as chapter 9 (on C2).

and *"stands on behalf of the parents"* (mothers, school 1, district C). By this respondents refer to SCs responsibility in enforcing parents' priorities. As discussed in the sections on communication, and on engagement, it is however unclear the extent to which this occurs in practice and whether parents' do feed into the school committee agenda.

To monitor pupils' attendance, SCs collaborate with teachers, head teachers and community leaders to make follow ups on those pupils who tend to miss school more often. This is usually done by sending pupils found wandering around the village during school time to school. Additionally, schools perceive the SCs responsibilities to involve sensitising parents on the importance of education in order to improve pupils' attendance. SC members refer to going to individual parents' houses to encourage them to enrol/bring their children to school, and to speak to the parents and the wider community during parents' and village meetings. Lastly, as discussed in Chapter 9 in Volume II on SLM, the role of the village officer in enforcing pupil's attendance appear to have increased, and SCs say they report parents to the village officer if they refuse to follow SCs' advice.

Lastly, SC members in all schools comment that they check on teachers' attendance and commitment by regularly supervising teachers' activities and behaviour in school. When issues emerge, the committee brings it up at school meetings, and if the problem persists SCs bring it to the attention of community authorities. However, in the majority of schools other respondents do not validate this, indicating that whilst SCs themselves appear aware that this is part of their responsibilities they do not necessarily fulfil this. In two schools, schools say that SC members do at times check in on classes, though this appears to be ad hoc rather than structured monitoring.

However, data show that in some cases the actual functioning of SCs is scarce. **Some SCs perceive lack of cooperation at different levels (parents/community and district/government level) to pose constraints to their overall capacity and motivation.** Several SCs find it difficult to arrange meetings as members do not want to show up, due to not receiving incentives. In one school, the SC took the time to meet following the FGD, as that was the first time they had had members show up when called for a meeting (school 2, district A). In some cases, respondents state that the lack of financial compensation can demotivate members of the committee. As one teacher puts it: *"there are some [members] who do not attend meetings, because there are no allowances. [This] demoralises the committee"* (school 3, district A). In one case, the new head teacher has decided to give money to the committee in order to encourage members to come to meetings (school committee, school 1, district A).

Besides wanting monetary incentives, all SCs, as well as other respondents, note the need for more training for SCs to be able to perform their duties successfully (as many SCs members have received little to no training). As a teacher in school 3, district C puts is:

> *"The selection of school committee members was done but [did not start until recently] to do their activities…after the selection they were told that there [was going to] be a seminar but no seminar [was] conducted. They started doing work because there was money coming in for school projects and sports items. So, the head teacher had to tell [them] their responsibilities".*

This may reference the head teacher cascade model, but indicates that SC members themselves may not always be taking part in training (as discussed above). It does, however, show that SC members may be expecting training to occur, and for their 'roles and responsibilities' to only start once they have been told what to do. **In many cases, the head teacher appears key to the effectiveness of SCs' work.** SCs in communities where the head teacher *"involves the committees in anything that happen within the school"* (mothers, school 3, district A) appear to function better due to better communication and increased motivation to engage with the school. It also appears that head teachers in these schools instruct SCs of what to do, as such fulfilling the expectation of SCs that someone needs to inform them of their responsibilities more fully.

One of the responsibilities of SCs emerging from EQUIP-T is the responsibility to manage the EQUIP-T grants. As discussed, data shows that SCs together with head teachers are involved in managing school budgets. As such, SCs are responsible for the EQUIP-T grants schools receive. What seems to have changed in the last two years is that SCs now are perceived as "*the school owner, the school cannot buy anything until they have received consent [from the SC]*" (teachers, school 1, district A). For example, as a teacher in school 3, district C explained:

> "…*when you do something you have to involve [the] school committee. For example, [when] we received the money [from EQUIP-T] [we] informed the committee [what] amount [is] needed [for] school projects and sports items. [For example}…we want to build [a] toilet for the teachers [so we asked the SC]: what do you [think]? The committee discussed [it] and said: 'this is a good idea'".*

Thus, SCs have to approve or check all the resources and money that schools receive before they can actually use them. SCs say that there are certain guidelines, and that they have to use grants to address concrete school needs, such as getting new desks, building toilets, or providing sports facilities. Discussions and proposals on how to invest the money occur during SCs' meetings. However, relating to the above discussion that SCs struggle to find the time to meet due to lack of motivation and morale, this may affect the implementation of this internal structure. However, it is not always clear whether stakeholders are referring to EQUIP-T grants or government money or both of them. SCs also report to manage capitation money that comes from the government, and it is not always clearly distinguished what EQUIP-T money is spent on vis-à-vis capitation grants. This thus poses difficulties in determining the relative effect of the EQUIP-T grants. Yet, **the qualitative data suggests that EQUIP-T has given SCs more power to manage school budget and supervise school resources, as the budget increases with the EQUIP-T grant.**

There is some inconsistency in the data regarding how SCs interact with communities. For example, whereas parents and teachers in school 2 of district A report that their interaction with the SC is limited to school meetings and perceive that they are hardly involved in the committee's decision making and planning processes, the SC and community leader in this school claim that SC engages with parents and the broader community before making any decisions. This inconsistency may be explained either due to SCs knowing they are expected to engage with the community (and thus for there to be a social desirability element) or that the parents spoken to as part of the FGD were not widely engaged. Regardless, **there is a general perception that SCs are more active and engaged both within the school and between school and community than in previous years**. One of the main effects attributed to this interaction and engagement seems to be an overall improvement in pupils' attendance in the three districts. Notably though, SCs and other stakeholders, especially parents and community leaders, tend to attribute improvement in SCs' attitude and work to the new head teachers' ability to supervise and involve the committee. EQUIP-T is not mentioned unless probed (when it comes to the effect on SC work) and the perceived contribution made by EQUIP-T tends to be limited to the grants. Still, through grants, SDPs and head teacher involvement **EQUIP-T appears to play a role in the changes, in spite of respondents perceiving other factors such as a change in head teacher as having a more direct impact.**

Consequently, the qualitative research indicates that the assumptions that training is effective, all SCs have attended, and they are now aware of their roles and responsibilities and capable of perform their duties do not hold for all the cases at midline. This appears related to several factors. First, an uneven management and implementation of EQUIP-T training, across and within different districts, appears to affect the extent of change between the way SCs function now and how they used to function before EQUIP-T. Second, the expectations of SCs regarding the training they ought to achieve seems inconsistent with the planned programme training. Third, the motivation and capacity of SCs to perform their duties vary significantly across cases, and seems highly dependent on the head teacher's role in involving and motivating the SC. As such, though data

indicates increased awareness around roles and responsibilities in SCs who have received seminars, the activity of SCs across schools appears low.

11.2 Communities' capacity-building: PTPs

One of the key outputs of EQUIP-T under component four, is the establishment of PTPs to enhance parents' participation and capacity to support school improvement and, ultimately, feel able to hold the school to account. The qualitative midline thus conducted FGDs with mothers and fathers in each case school to explore their understanding and perceptions of the school and school leadership and management.

The qualitative findings show that 8 out of 9 schools have elected PTPs. In the case school that has not yet elected a PTP, respondents show no awareness of what a PTP is. It is difficult to know whether this is due to a lack of implementation externally, or internally, though neither teachers nor parents are aware of the PTP in this school.

In the 8 schools that report having established PTPs, the election process is fairly standardised. 14 parents and 7 teachers (class teachers) sit on the PTP, with parents electing one father and one mother from each class during a parents meeting. As such, **PTPs appear gender balanced, with parent representatives having children in the class in which they represent together with the class teacher from that standard.** In one of the schools a father say he received an EQUIP-T seminar on how to set up a PTP (parents in other schools did not mention seminars, though some head teachers had attended seminars about PTPs):

> *"When we were given seminar on this EQUIP-T programme we were told that every class should have two representatives that is from grade one to grade seven to make a total of fourteen parents with seven teachers and become twenty one members"* (Head teacher, school 3, district A).

However, it seems that not all the case study schools align with the intended PTP make-up. In one case, 2 representatives of the pre-primary school are also part of the PTP, together with 2 teachers per class (school 1, district C) and in another school, there are no representatives for standard I (school 3, district B). There is some further confusion in two other schools (school 2, district B; school 3, district A) where some respondents refer to the PTP only having 14 members, and as such only one parent representative per class. However, it is unclear whether this is the case, or a misunderstanding on the part of the respondent. Nonetheless, there is a clear confusion around the role of the PTP, as well as who sits on it, as discussed below. Additionally, the role and involvement of the head teacher on the PTP varies, with the head teacher often taking up roles as secretaries or supervisors. In some of the schools visited the head teacher is also a class teacher, and as such holds a position on the PTP.

11.2.1 PTPs: roles and responsibilities

Overall, schools show different degrees of understanding and awareness of PTPs' existence, composition and intended roles and responsibilities. This seems to be mainly due to the fact that **the majority of PTPs, although established, are not active**. In several of the 8 case schools with an established PTP, respondents only acknowledge this after explicit prompting. It appears that though these cases elected PTPs – only some schools have PTPs still in place and that are active. For example, in school 2, district A, the head teacher said that though they had elected the PTP, nothing happened since – and she had never seen the parents since that moment.

Interestingly, many parents and teachers, unless they are part of the PTP, are not aware of the PTP existing. It appears that only parents who were present when electing the PTP members are aware of its existence (thus including elected members). Since respondents state that nothing has happened since the establishment, other parents and school respondents are less likely to be

aware of the PTP. As one school committee member puts it: "*it is difficult [to] remember something you initiated and [then] left without implementation*" (school 1, district C). In the FGDs this was evident, as only those FGDs that happened to include a PTP member in the sample could gather data on the PTP. This meant that although head teachers and school committees may mention PTPs, and PTP members themselves may state that they are active, it appears from these FGDs that the wider parent community are not aware. Even in cases where parents may know of the establishment of PTPs, they do not perceive it to be active. As fathers in one school discuss:

> "*Interviewer: does the school have a PTP?*
> *Participants: what is that?*
> *Interviewer: a PTP, I mean a parent-teacher partnership. Does such a group exist here?*
> *Participants: no*
> *Interviewer: hasn't anyone ever heard of the PTP?*
> *Participants: no, it hasn't happened here*
> *Note taker: haven't you ever heard of representative parents from each class with teachers?*
> *Participant 3: yes, they are there…it exists but it does nothing*
> *Interviewer: what does the PTP do at the school?*
> *Participants: we don't know*
> *Participant 3: we don't know about that because it wasn't given priority*"
> (School 3, district B).

Additionally, respondents in schools where a PTP is more active, or where head teachers and teachers are aware of PTPs, perceive it as a sub-section of the school committee. Schools, parents and community leaders describe the SC as the SLM body representing the community, and for the PTP to sit within that already established and acknowledged committee. As a father explains: "*the PTP is found in the school committee. It has the chairperson, the secretary and the members*" (school 1, district B).

There appears to be an overall blurred perception of the difference between PTPs and SCs, and respondents, including SCs themselves find it difficult to distinguish and separate the bodies. One SC say this affects the effectiveness of the PTP, as "*we haven't yet sat and discussed together with the PTP, because after the election we have not [got] directives on their responsibilities*" (school 1, district C). Moreover, the same school refers to issues with regards to authority, where the PTP see themselves as part of the SC, and as such to have SLM power, whilst the rest of SLM and teachers see them as working for the school "*the main constrain is that they think they are [the] SC, which is not true. They are working without any training, so they are working according to HT instructions*" (Head teacher, school 1, district C). **This notion that the SC is an SLM actor, whilst the PTP is undefined, appears a challenge for PTPs.** For example, one teacher report that parents challenge the PTPs authority instead of collaborating with them on issues such as pupil absenteeism (school 1, district B). However, it is unclear whether this relates to parent's general view on education rather than a lack of recognition of the PTP as an independent authority. However, **if one of the main goals of PTPs is to close the gap between parents and teachers, and enhance parents' participation in school matters, this lack of definition around where the PTP sits in regards to other school and community actors might have an adverse effect.**

Moreover, stakeholders perceive there to be a link between PTP inactivity and the lack of capacity building available for PTP members. Notably, **respondents expect EQUIP-T to provide training for PTPs**, which contradicts the programme's assumption that PTPs will be self-organised with minimal training. In fact, some respondents legitimise PTPs' inactivity due to the absence of training, with one school committee member stating: "*they haven't done anything so far. And it's not their fault because they have been told that one day they will [be] called [and] assigned their duties and responsibilities*" (school 3, district B). **PTPs thus often appear to be waiting for training to occur before implementing anything, which explains the lack of activity post-election.**

As a result, **the general perception amongst respondents is that PTPs are mostly not aware of their roles and responsibilities**, due to not having received training. Still, where PTP members, school committees, teachers and head teachers have heard of PTPs they perceive the following as its main role and responsibilities:

- Monitoring the classroom (supporting the teachers to identify and address pupils' issues, and checking on pupils' and teachers' absenteeism, acting as a bridge between families and the school);

- Overseeing all EQUIP-T class-related projects; and

- Managing EQUIP-T budget together with the School Committee.

As discussed, these activities appear to rarely occur, but **in some schools PTPs show some activity in spite of not receiving training**. In school 1, district C and school 3, district A; head teachers, teachers and PTP members report that PTPs rely on class observations and networking with relevant stakeholders (SCs, teachers and head teachers) to address issues that arise at class level. Parents mention coming to school to check that class is happening, but that this only happens now and then.

The role of the head teacher in the PTP appears to affect the roles and responsibilities of the body. In schools where PTPs are somewhat active, but have not received training, the head teachers appear to act as 'supervisors' of the PTP, defining their role and responsibilities. For example, in school 3, district A, the head teacher has assigned the responsibility of health education to members of the PTP, where mothers and fathers come and speak to girls and boys respectively about puberty and personal hygiene (head teacher, school 3, district A). When asked why the he assigns this responsibility to the PTP, the head teacher explained: "w*e decided [this] because SMC is powerful [and] has many responsibilities"* (school 3, district A). As such, this school defines the PTPs role and responsibility in relation to that of the SC, so as not to overlap. **With the head teacher directing the roles and responsibilities of the PTP, and doing so taking into consideration the power balance between the PTP and the SC, PTPs may lose some of the agency and initiative envisioned through EQUIP-T's aim of empowering parents and making them active participants in the running of the school.**

With regards to school budgets, **stakeholders report that EQUIP-T gives 550,000 shillings to the school, of which 100,000 is managed by the PTP and the rest by the SC**. Though the programme states the need to apply for PTP grants, the majority of stakeholders across the three districts say that PTPs have received money from EQUIP-T without applying for it. However, this does not necessarily mean that schools have not since previous head teachers may have supported applications, or PTP members who were not part of the qualitative sample led them. The head teacher in school 2, district B is the only one referring to actively filling in a form for EQUIP-T together with PTP members, with the purpose of *"why we were asking for that sum of money and how we would use it".*

Interestingly, **schools appear to just have received their first PTP grant**, which may explain why PTPs in the majority of schools have not been active yet. In school 1, district A, the PTP were having their first meeting on the day when the qualitative team were there[47]. When the school receive EQUIP-T funds (including the PTP grant), PTPs and SCs sit together to decide where to allocate some of the money**.** It seems that the PTP can use the 100,000 shillings for smaller projects and expenditures while the SC uses the rest for bigger school project such as buying cement to build teachers' toilet, to address major schools' structural problems or to implement IGAs, such as the cultivation of sisal. **There appears to be a lack of clarity and guidance around the purpose of the PTP grant**, with PTPs spending on anything from glucose to sports

---

[47] It is important to consider here that the WEC had told this school in advance that the qualitative team were coming. It is not clear whether the PTP would have met on that day had it not been for the team visiting.

equipment or chicken. As PTPs have not received training, and in many cases appear to have had little involvement with the school prior to receiving the grant, it is questionable the extent to which PTPs are best placed to decide how this money is spent. The qualitative data is not able to tease out the extent to which PTPs have agency in making this decision, or whether SCs and head teachers play a major role in this, as the sample did not include PTP members specifically.

Lastly, apart from lack of training and unawareness of roles and responsibilities, **PTPs activity appears affected by the capacity or motivation members have to fulfil their roles[48].** Some head teachers and parents state that the lack of allowances represents the main obstacle to accomplishing PTP's duties and is likely to affect sustainability. On the contrary, head teachers in school 2, district B and school 3, district A consider PTPs to be a sustainable component of EQUIP-T precisely because there are no allowances/incentives for PTP members to take part. Otherwise, once allowances end, activities end and thus if PTP members can become active without allowances their activities are likely to continue after the programme ends. However, in the case study districts, pupils' families are poor and often pastoralists. Being a PTP member is a voluntary position, and parents thus endure an opportunity cost each time they take part in PTP- or school activities[49]. Consequently, parents prefer to stay home and take care of their households or IGAs instead of spending their time dealing with school issues. As fathers in school 1, district B discussed:

> *"Father 4: some of the parents find it difficult because they are unpaid. They prefer to stay at home instead of doing that [PTP] work.*
>
> *Father 2: the other thing is transport challenges…they have to move around the community to sensitise other parents to make sure that children attend school"*

However, **respondents are generally positive to the idea of PTPs**. Where operative, schools perceive PTPs have brought positive changes by keeping parents involved in education issues and, hence, raising their awareness and empowering them. However, whilst the election of PTPs symbolically does this, the lack of activity of PTPs in most schools means that one must be careful regarding these statements. **The perception may be that change occurs, or rather could occur, through establishing PTPs, but the qualitative data at midline is not able to provide concrete examples where PTPs have contributed to the above perceived changes.** Yet, in schools where there is PTP activity (school 3, district A; school 1 district B; school 1, district C) respondents perceive PTPs to contribute to improved relations between schools and communities and to a decrease in pupil absenteeism as PTPs help SCs by speaking to parents and sending children who are wandering around during school hours to attend class.

11.3   Community-led school needs assessment

Another key output of EQUIP-T is the community-led school needs assessment, aimed at fostering community's involvement in schools. According to the programme, this needs assessment is meant to be led by a community facilitator (CF) who is trained by a CSO. The CF then conducts the school needs assessment together with a village task force (VTF) and develop and action plan that feeds into the school development plan.

11.3.1 Community Facilitators and Village Task Force

The CF should be a member of the community whom, once trained on roles and responsibilities, is responsible for conducting the community-led school needs assessment together with VTF.

---

[48] It is important to note that the qualitative research did not purposively sample PTP members. As such these are the perceptions and views of other respondents (as well as a few PTP members, who ended up part of the parents FGDs).
[49] See further discussion in Section 11.4.3 on parent's engagement in schools.

Even prompted, almost all stakeholders are unaware of CF's role and responsibilities and in most of the cases they say they do not know who the CF is or that their community does not have one. The only exceptions are informants in school 1 in district C and in school 3 district B, where respondents say they selected a CF (though it is not clear whether through parents or village meetings). In school 3 district B, the community facilitator was a member of the SC and as such part of the qualitative FGD in that school. He said: *"I got [a] seminar…I waited for the trainer who came [to the] school and I was put together with more people whom I will be working with and they are called [the] village task force. […]"* (CF/SC member, school 3, district B). In this case, he was told to attend training at the ward level but he himself does not know who chose him for this role, whether the school or the village government. He also could not clearly communicate his role and responsibilities as CF, and no further implementation appear to have taken place.

In school 1, district C it is unclear whether respondents are referring to the CF or another community actor such as a school guardian. The community leader (CL) in this school perceives the CF as a school guardian whose role it is to educate the community on school development issues (community leader, school 1, district C). After receiving training at ward-level, he has started to regularly come to school to supervise that pupils get porridge and equal treatment, that money "*brought by the organizations*" (school committee, school 1, district C) are well allocated and effectively used, and that teachers are teaching.

In school 2, district B, the head teacher said he had been told about a CF in an EQUIP-T seminar but that he had not heard more about it since. His understanding was that the CF needed to be *"a man…skilful…should have secondary education"* and for the responsibilities to include supporting teacher's activities in the community.

However, whilst the role of the CF is to represent the community and facilitate their involvement in the school through the conduction of community-led school needs assessments, it clearly emerges from the data that this is not occurring. In fact, the community instead perceives the CL to be the bridge between the school and the community.

Similarly, stakeholders across the three districts do not know about a VTF. Apart from the CF in school 3, district B, there is no mention of a EQUIP-T specific VTF. It is important to mention that all communities refer to other VTFs, already in place before EQUIP-T. Each community have various VTFs that focus on infrastructure, health and other issues in the community. However, an education specific VTF does not appear to be in place, though the infrastructural responsibilities of VTFs also covers schools. Overall, there seems to be an issue of definitions, with stakeholders' understanding of each actor's role and responsibilities lacking clarity. It seems that respondents acknowledge and are aware of those bodies or roles (SCs and Community Leaders) that were in place or had responsibilities even before EQUIP, while there is a lack of understanding of the roles and responsibilities of EQUIP-T established actors (PTPs, CFs and VTFs).

11.3.2 Community-led School Needs Assessment and SDPs

Those SLM actors and parents who said they knew about community-led assessments perceive that "*it is done during the village meetings where a community member can stand up and talk about the school weaknesses and other challenges so the meeting decides to discuss how to solve those problems*" (school committee, school 3, district B). As such, **communities appear to confuse community-led school needs assessment with village meetings,** and state these to be where the school committee or head teacher informs the community of school needs where after the community debates over the issues to be solved, and evaluates whether the community has the capabilities to address those issues and "*advice what they can support or devise possible solutions*" (head teacher, school 2, district B). Thus, though respondents may say they have conducted a community-led school needs assessment, this appears blurred with any process in which the community is involved.

However, the above perceptions may refer to the school development plan process[50], but without an underlying community-led school needs assessment, instead coming from the school to the community rather than the other way around. Data show that **in some cases parents, when asked about community-led needs assessment, refer to the SDP that SCs conduct, and perceive that the community is hardly involved in any assessment processes.** As one mother says: *"personally, what I see is that the school committee meets and discusses everything about the school. If you are not a member of the school committee is difficult to know"* (mothers, school 1, district B). Other cases echo this notion of a top-down assessment. For example in school 2, district B both the SC and fathers mention the SC to always do the assessment and that *"the community, is always represented by the leaders who sit and see what the school requires"* (school committee, school 2, district B).

Thus, overall, **respondents appear unaware of what a community led-school needs assessment is**, and whether their community has done one. The qualitative research shows that in some of the communities, especially if district B and C, even when prompted, parents state they have never gathered together with the community and the teachers to list all the school's needs and weaknesses and discuss on ways to address the needs. **As CFs and VTFs in most cases do not appear to have been selected, and in cases where they have do not seem to have enacted their roles and responsibilities, it is likely that this component has not been implemented as intended.**

11.4 Communication, Participation and Accountability

Because of EQUIP-T inputs, the expected outcome is for communication to improve between schools and communities. Following better information and transparency, communities and parents will be more actively involved and aware of what is going on in the schools.

11.4.1 Notice boards, communication and transparency

To encourage increased communication between schools and communities, and higher transparency, EQUIP-T has provided noticeboards to schools. **The qualitative research found 8 out of 9 schools to have a noticeboard, and respondents frequently referred to the noticeboard as one of the resources the school has received from the programme.** However, only 3 of the schools visited (all in district A) displayed the noticeboard outside on the school building, whilst 2 schools had them inside the teachers' office, 2 inside the head teachers office and in 1 school it is not clear where the noticeboard was located, but it was referenced by respondents. Interestingly, all schools in district A had their noticeboards on display. In at least one of these schools, potentially two, it seemed like the noticeboards are not normally hanging outside the school building. In school 3, district A the qualitative research team found groups of pupils standing in front of the noticeboard, acting like it was a novelty. The team ask various pupils who all said that the noticeboard was hung there the day before.

The noticeboards observed by the qualitative team displayed the names of school committee members, PTP members, the school budget as well as the number of pupils currently studying in the school. In one school, the noticeboard was used to display class-rankings (i.e. how pupils had performed in exams). **Parents and community members did however not refer to the noticeboard as a source of information.** Considering that the majority of schools keep the notice boards indoors, inside offices, it is unlikely to fulfil its intended objective of increasing information available to parents, and wider transparency. Additionally, **as parents rarely come to school (discussed further below), they are less likely to see the noticeboard as a means of finding out what is going on in the school.**

---

[50] See further discussion in section 11.4.3 on school engagement and Chapter 9 on the SDP in SLM.

Instead, **a main way of communication appears to be village meetings**. Schools across different districts report that, firstly, SCs, head teachers and teachers meet together at school to discuss about school's matters. Then, through CLs first, and VEOs afterwards, they involve the community for issues that require their engagement (i.e. school's infrastructure, management of school budget, and teachers' houses). During village meetings, community members "*also get a chance to give their opinions on how to solve [a] challenge*" (community leader, school 1, district C). **CLs and VEOs act as a bridge between schools and communities, facilitating communication.**

To inform the community of meetings, and to gather members together, SLM actors use different methods across case schools. In some communities, CLs or VEOs blow whistles, or "*use drums and pass around the village to tell them* [community members] *the time that they are being needed at school on a certain time and day*" (school committee, school 1, district B). Alternatively, schools inform communities of meetings through letters and/or through pupils. For matters that need to be as widespread as possible (such as pupils' registration for enrolment), some schools report that they recently have started to place information around the communities. For example, teachers in school 2 district B say they prepare and stick leaflets:

> "*…on the wall or notice board of the village government office. We mostly target places where more people gather, for example in school we stick the announcement on the trees so that those who use [the] football ground can see and inform other people in the community*".

Similarly, teachers say the school committee in this area use churches and mosques to inform communities of school registration, which is something that did not previously happen. As one teacher says: "*we have never heard announcing it* [information on registration] *in the churches and mosque, but nowadays when we attend the church especially in December we can hear them announcing*" (school 2, district B). **Respondents in several schools refer to communication to be better if information is placed within the communities themselves, rather than at the school, since it makes it more accessible for parents and community members.** As discussed above, the EQUIP-T input of providing noticeboards to schools may thus not be that effective if displayed at school, but rather information needs to reach communities themselves (at least in cases where the school is far from communities).

In those cases where it is the community that needs to communicate with the school (besides village meetings and the resort to CLs and VEOs), some teachers say that they can write letter to the head teacher (school 2, district C). However, parents and community members themselves say they will use community actors such as community leaders, VEOs or school committees to communicate with schools. Alternatively parents say they will go to the school to see the head teacher if there is an issue, though mainly to do so when the school asks them to come. No parents reported to have sent letters to the school.

The majority of head teachers express frustration identifying the best ways to communicate with parents and communities in order to increase their involvement. Head teachers find it challenging to get parents to attend meetings at school, and thus find it difficult to be as transparent and inclusive as they are taught to be (in cases where they have attended SLM training). In this manner, **head teachers perceive their best avenue for including the community to be through the school committee and community leader**, but at times face difficulties in managing these relations as even SCs and community leaders are not always motivated to actively involve themselves in school development.

The qualitative research found that the main means of communication between parents and schools to be through letters or oral messages that pupils have to deliver to their parents or relatives. If the pupil is absent, teachers give the letter to another student living nearby. Various FGDs with pupils found that pupils feel responsible for delivering these messages or letters. Although the majority of them seem to accomplish this, some pupils report that they "*tear and throw them* [the letters]*"* (children school 1, district B). Therefore, this method of communication

appears to not always be reliable. If parents are not able to read Kiswahili, pupils have to read for them.

Another new way to communicate with parents is through phone calls. Teachers report that *"if you have the phone number, you just call them [to] 'come to school we need you'"* (teachers, school 1, district C). One head teacher said he has drawn up a register for parents, so that they are better able to contact parents through either notes or phone calls (school 2, district B). He found that parents are less likely to respond to notes than they are to phone calls, but find it to be logistically challenging to phone each parent for meetings. Instead, he introduced a 'RSVP' slip on the bottom of the note, which helps him identify if parents have received the notes or not. Respondents report that when parents do not have a phone, schools resort to letters to convey messages.

Both school and community respondents perceive that overall, communications between parents/communities and schools have improved in the last two years. Both parents, SLM actors and teachers report that schools now contact parents not only to inform them on the forthcoming meeting or on examinations' results at the end of the year, but also to discuss more pupil issues (absenteeism, behaviour at school, commitment to studying) than in previous years.

However, respondents do not directly reference the role of EQUIP-T in improving communication and transparency. Still, **head teachers, SCs and community leaders refer to there being more transparency now than before EQUIP-T, attributing this to an open dialogue between schools and communities through discussions around the SDP and budget management**. The community leader in school 1, district A explains the increased transparency a result of the openness around the SDP:

> *"Things are open, like [the] School Development Plan…community members get involved to listen, see and even given a chance to advice [on the] budget,… nowadays the school and community plan together on what to be done [with] money"* (community leader, school 1, district C).

As discussed in chapter 9 in Volume II on SLM, **respondents perceive the head teacher as key in ensuring good relations between the school and the wider community**. This is mainly in reference to the SDP, where communities are included as an active part in shaping the school agenda. Communities state, that though they do not need to be involved in decision-making within the school, they prefer to know what is going on. Thus, parents and communities appreciate it when the head teacher (i.e. the school) is transparent. **As such, it appears that EQUIP-T positively effects transparency through the SDP, though community respondents are not necessarily aware of EQUIP-T's role in this.**

11.4.2 Relationship between school and parents/community

Teachers across all case study schools blame parents for pupils' absenteeism and clearly mention that parents' attitude towards education is one of the main challenges between teachers and parents. As one teacher puts it:

> *"…you find children are coming to report: 'my father asked me not to come to school', 'mother asked me to not attend school'. This creates a conflict between parents and teachers. We believe that children do not tell lies, and when they come to school they say: 'teacher don't punish me, I like school but my father said we have to go to weed the tobacco farm or apply insecticide or do this and this'"* (school 2, district B).

As the above quote illustrates, **teachers do acknowledge that there is an economic element to parents keeping their children from attending school.** Many families are poor and need their children to contribute to the household. Teachers report that the lack of economic returns to education reinforce the reluctance that parents have in bringing their children to school. This is

particular in reference to pastoralist communities, where teachers and other school actors, as well as community leaders report parents to consider education as superfluous, and in fact, a barrier to their pastoralist way of life. One head teacher explained that pastoralist families have always managed to provide for their household through farming and grazing, without receiving education themselves. As such, the perception is that parents in these communities do not see the importance of getting an education and therefore are not motivated to send their children to school. As one teacher explains:

> *"When comparing educated and non-educated [people, you see], in [a] pastoralist society, if you have cows and your children do not go to school they will benefit from the cows. Hence they do not see the meaning* [of education]. *Those [who were] educated have not been employed. Therefore they will be the same as the ones who did not go school. [So parents] start comparing 'we spend all the money and get nothing'…This makes [it] hard for them to change their perception and practices"* (school 3, district C).

To this extent, **teachers see pastoralism as largely affecting parent's wider attitudes towards education.** Additionally, **schools perceive 'pastoralism' to affect pupil's attendance and poor performance at school.** Schools perceptions of parents attitudes towards education was generally more negative in those case schools located in pastoralist communities, with uneducated parents from pastoralist communities often viewed as responsible for pupils' low attendance and poor performance at school. Interestingly, some parents support this notion, saying that:

> *"A tribe like [ours], Sukuma, is known to not like education. They don't insist on education, they just do it to fulfil their responsibilities. Parents don't want their children to go to school, they are telling them to answer wrongly in the examinations to show they don't know anything and won't move forward"* (mothers, school 1, district C).

However, when asked what they expect from the school, all parents in FGDs agree that they want their children to get a good quality education to be able to pass exams and get a good job. Baseline showed that teachers would often see parents as uneducated and 'backwards' and for this perception to be a source of resentment for both sides, negatively affecting relations. This was not brought up at midline, but rather parents and other respondents report that parents and **communities are becoming increasingly aware of the importance of education.** Whilst teachers still refer to lack of awareness of the importance of education as a challenge for teaching, parents would acknowledge this as well but state that it is improving. Teachers and SLM actors echo this and refer to EQUIP-T's focus on community involvement to help improve parent's attitudes to education, and in turn the relations between schools and communities: *"the education provided by EQUIP-T to parents, teachers and pupils have brought changes by encouraging parents to get involved in education issues"* (teachers, school 1, district C).

These references to the increased awareness of the importance of education is however somewhat in contrast to other statements from pupils' and teachers that parents keep their children from school to do farming and other IGAs. Still, it appears that **parents may positively view the idea of education, though not always seeing it as feasible within their own realities**. Yet, respondents state that the economic returns to education have played an important role in changing parents' view of education as *"most of parents did not know the importance of education, now they have seen some people who attended primary school here, that now are at University of Dodoma (UDOM)"* (community leader, school 1, district A). It seems that being educated and seeing positive educational outcomes in terms of better job opportunities encourage parents to send their children to school. One head teacher said he uses himself as an example to show communities how education can provide you with new opportunities, saying that: "*I counselled them [on the importance of education]. As I also come from a pastoralist family, but now I am at this level. Therefore things are going well"* (head teacher, school 1, district A). Relatedly, **parents report that seeing their children perform well at school works as motivator to have their children educated.**

The extent to which teachers are a part of the community in which they live seem to affect the relationship between parents and teachers (as discussed in Chapter 8 Volume II on teachers). In those cases where teachers are seen as part of community (such as in school 2, district 1), the relationship between is perceived as good. Parent's say they do not feel encouraged to engage with teachers, as they only see them during meetings, and as such for teachers to not be interested in getting to know parents. Teachers themselves feel that parents want them to get involved, but at their terms, whilst teachers feel they do not always fit into the communities in which they are placed: "i*f you don't mix with them in their gathering and play gambling or wear a nice cloth they regarded you as someone who just show off"* (teachers, school 2, district B). This proximity, or 'mixing' and its effect on relations, seems related to whether schools have teacher houses or not, and as such whether teachers stay in the communities after school closes or if they commute to the city. However, some respondents report conflicts between, in particular, female teachers and communities, as their lifestyles are not seen as proper, and out with community norms (community leader, school 1, district C).

Moreover, **though communication may have increased this does not necessarily result in an improved relationship between parents and teachers.** For example, in school 1, district A, respondents continuously refer to improved communication, but teachers say: *"there are no changes in the relationship with the community, because they hate teachers"* (school 1, district A). Therefore, the existence of a dialogue between schools and communities not always implies good relationships or participation. As discussed above, this seems related to contrasting ideas of what is best/proper for the school, the community and for pupils. Some teachers on the other hand tend to see parents as obstacles to changes they want to make. As one teacher explains:

> *"It reached a time when we set out road bumps, because cars are passing here with high speed, but some parents removed all road bumps that we set and said: 'what is the meaning of doing that, we have been without road bumps for so many years, they think they know better than us' (Teachers, school 2, district B)*

Contrasting this, parents at times perceive teachers to always *"give directives, there is this and this and this"* (mothers, school 3, district A), and in this sense look down on them slightly, as exemplified by the above quote. **As teachers come to rural schools in communities different from their own, this clash is likely to occur, with teachers seen as imposing unwanted changes rather than being part of the community and its development.**

This relates further to the issue of corporal punishment, highlighted at baseline as a key source for conflict between parents and teachers. At midline, the majority of respondents would bring this up as a conflict that was in the past, and for schools and communities to have reached, or started dialogue around reaching, agreements regarding punishments for pupils. As discussed by fathers in school 2, district B: *"there have been changes because the issue of the parent to become angry whenever their children are being punished has been taken care of by sitting with them in meetings and talking about this"*.

However, as discussed throughout this report, there is a potential element of social desirability bias here, with teachers and head teachers stating corporal punishment as less of an issue now when they have been taught to punish pupils less. Still, **respondents generally perceive corporal punishment as less of a source for conflict now than in previous years.** For example in school 1, district C parents said punishments or 'cruelty' had decreased after the head teacher received training. Rather than parents disagreeing with teachers punishing their children, the conflict appears to be around when teachers punish pupils too severely. Parents agree about giving pupils 'normal punishments'.

As such, some schools and communities have solved this issue by discussing how many times a teacher can use a cane on a pupil. For example in school 3, district A, teachers are not allowed to use the cane more than 2-8 times on a pupil, depending on what they are being punished for.

Additionally, respondents refer to better communication as decreasing conflicts regarding this – and for head teachers to be better at speaking with parents if they have issues, instead of parents going directly to teachers and shouting at them. **As discussed in chapter 9 in Volume II on SLM, the relationship with teachers and the community was widely highlighted as a responsibility of the head teacher, by both head teachers, teachers and the wider community (parents, school committees, and community leaders).** Respondents perceive the training that head teachers receive (in the cases where they do), or even teacher's training regarding punishments, to have led to schools working to more closely consult parents on issues that may cause conflict (such as corporal punishment).

Lastly, **another area causing tension between schools and communities is the issue of early marriages.** The majority of the communities visited still encourage early marriages, especially for girls, though the 'appropriate ages' (ranging from 12-16) differ between cases. Schools say they face difficulties retaining girls in school, with parents at times pulling the children from school and pretending that they sent their child to live with a relative or that they are ill. The head teacher in school 2, district C says the school committee has the responsibility of addressing this issue, and that they have held village meetings to discuss it with the community. Nonetheless, this is still a delicate terrain, and an issue that is prevalent across the nine case schools. **As discussed in chapter 1 on pupil learning in Volume I, though girls are less likely to be married before standard 7, this can still affect both girls' motivation to learn, and parents motivation to send them to school if the aim is to marry them regardless.**

11.4.3 Community engagement and involvement in schools

The involvement of communities in education appears to have improved, although it seems to be a slow process. Respondents say that community members attend village meetings more (regularly), and respond more positively to requests from schools (such as helping to build or repair school infrastructure, toilets, teachers' houses etc.). However, issues of communities' priorities and value of education (discussed above) are a crucial when it comes to dealing with their involvement in schools. Teachers in school 3, district C, referring to community's attendance at meetings, report that: *"if you tell them they have to come at 9am you can expect them at 3pm. Those who are aware and motivated in school matters will come at 1pm".* In other cases SLM actors say that there have been *"no changes because the community that surrounds the school are hard to understand"* (school committee, school 2, district C) to the point that some parents do not even want their children to be educated.

Two external factors seem to influence communities' participation and involvement. First, in some schools SLM actors have established a sanction (*masumule*) of 2,000 shillings that community members have to pay "*if you don't show up* [at meetings or when you are called], *the penalty is being given to that person"* (school committee, school 1, district B). This discourages members from missing meetings. Second, **respondents frequently cite perceptions around the new governmental law on free education as discouraging parents from being involved in school matters.** As a teacher in school 3, district C explains:

> *"You know this problem* [of community's lack of involvement]*, it is influenced by politics. We expect to [be able to] involve the community in making bricks and construction activities, but now when you try to ask the community to get involved in school matters they will tell you 'the government says the community is not allowed to do anything, the government itself will do it'".*

This finding is consistent across case schools. Schools thus state they find it difficult to involve communities after communities were told not to contribute to their children's education. Respondents tend to interpret 'community participation and involvement' mostly on a monetary/resource driven basis, with communities mostly involving themselves in schools either through payments or helping with construction. As such, schools perceive the law to have had an adverse effect on involvement (meaning monetary contributions in most cases). **Although this law**

does not seem to be an impediment to community's participation in all school's matters across case study sites, in some cases, as in school 3 of district C, it seems to work as a legitimisation for community's lack of involvement.

Still, as previously discussed, communities and parents get involved in school matters in various ways, through helping build classrooms or other school facilities, to showing up at meetings and collaborating with teachers to reduce pupil's absenteeism. In particular, the qualitative research found that parents appear more involved in monitoring pupil's education than previously.

Respondents perceive it to be parent's responsibility to monitor pupil's academic progress and teacher's commitment and attendance through checking children's exercise books. Some stakeholders acknowledge that there are limitations to this form of monitoring, in cases where parents are illiterate and are not able to check notebooks, if pupils copy others' exercise book or if teachers simply assign pupils tasks to copy without properly teaching to them, parents will not necessarily be able to assess the quality of teaching. Notably, even in cases where parents are aware of possible biases in judging teaching/learning only by checking notebooks, parents say they rarely (if ever) do follow ups. Moreover, parents feel they are not always able to monitor their pupil's education, but that teachers know better what goes on in the classroom. As such, parents in FGDs felt they were not able to comment on academic progress or the quality of teaching in the school.

The research found that the main reason behind parents' increased involvement lay in the relationship between parents and head/teachers. Where the relationship between teachers and parents is good, or have improved, parents feel more responsible for schools' development and pupils' education. As mothers in school 2, district C discussed:

"I see the involvement now is good but in previous years, they were not involving us, it is the first time I see it on my side […] The changes are there because if your child makes mistakes in school, the parent is being involved together with the teachers"

To this extent the relationship between parents and teachers function almost as a gatekeeper for parents' participation in their children's education. **Being involved with teachers makes parents feel more confident to actively participate in schools' life**: "*there are changes in which currently any time you want to come to school for the progress of the child there is no barrier, because we are very close to the teachers*" (fathers, school 1, district B). SLM actors in one school mention the PTP as a reason for why parents are now more involved (school 1, district B), as it directly creates a space for parents and teachers to meet. However, as the PTP in this school did not appear particularly active, these statements appear more theoretical, though the signal sent to the community through establishing a PTP may have contributed to this improved relationship.

Still, schools refer to the community component of EQUIP-T as the least functioning module, as parents rarely show up to meetings and do not come to schools when called. As discussed above, teachers cite lack of awareness as the key challenge for community involvement, and all respondents feel that EQUIP-T should focus on community awareness initiatives to help improve pupil's learning. Teachers in school 2, district B reference parents lack of involvement:

"Community involvement is very poor even when you call them to discuss about their children's progress in the school. You may find the child misbehaving in school, and when trying to call his/her parents so that you can help the child they don't show up. Also, at the end of each term parents are supposed to come to school to collect the exam reports for their children, to know he position of their children in the class, but still they don't show up and you may find only 20 parents show up out of 300".

As discussed in the section above on communication, **respondents in all case schools cite this lack of attendance in parents' meetings as a key challenge to involving the community.** Yet, the qualitative research found that in the majority of schools parents would show up when asked to

come for a FGD. In each school the team would call around 10 mothers and 10 fathers, with the hope that half of them would show. In some schools the team saw more parents show up than requested. For example, for mothers FGDs you would at times have the mother show up in place of the father for a household, as the father had to go to the farm, not wanting to be rude and not attend. Yet, this appeared related to one of two thing, either that the head teacher was on good terms with the community, or that respondents expected payment for attending a FGD (as other research teams had paid them). However, in some schools the research team struggled to get parents to attend, and had to go to the villages to find the sampled parents.

However, as with members of the PTP, parents feel they are not able to be as involved in school matters as may be expected of them. In all case schools, harsh life conditions limit communities and parents from being involved in education. Parents feel they are unable to come to the school or attend meetings as they have other priorities. As one father explains: *"the only constraint is hunger at home […] s you find yourself going for farming activities and not school matters"* (school 3, district C). Many parents live far away from the school, with some members of FGDs having walked for over an hour to attend the discussion. As such, there is not necessarily a link between low involvement and lack of interest of motivation, but rather that parents do not have the capacity in terms of time and resources to involve themselves in school matters.

11.4.4 Accountability

Many parents feel they are more empowered to hold school management and teachers to account for issues such as corporal punishment, child labour, IGAs and, to some extent, for budget management. They feel they can influence schools through recommending which farming activity to carry out, where to allocate money, and to help create a better environment for the pupils that is more conducive to learning. Most of them feel they can prevent teachers from beating pupils too harshly and from forcing children to do chores at teachers' houses. **Changes seem to be the result of effective and encouraged interactions between schools and communities that have taken place in the last two years through village and school meetings:**

> *"The community has an influence especially for the case of education, when they see that the education is deteriorating, they must ask the reasons for that to us, and we also have to take such information to required place and later give them feedback"* (school committee, school 2, district B).

However, some parents still feel that schools do not listen to their opinions, and that schools only consult parents on issues they know they will agree with. As such, parents feel that SLM actors do not share enough with them during school meetings, especially *"when they want to approve something that the community will not agree with, they do it on their own without calling community meeting"* (fathers, school 3, district A). **School budget management and money allocation seem to be the main area of debate where communities/parents do not feel involved enough, though as discussed above, respondents perceive this to have improved drastically since the recent introduction of SDPs.**

Additionally, parents state that they fear the repercussion if they tried to hold teachers to account for issues that directly concern their children's learning and treatment at school (such as attendance): *"it happened to my fellow parents. [They were] told that they do too many follow ups about their children in school, so we lose confident and get scared that teachers will not teach our children"* (mothers, school 3, district B). This notion that teachers will refuse to teach, or transfer from the school, is a prominent perception amongst parents, community leaders and school committees. All state that since they already struggle to motivate teachers to stay and teach,[51] they worry that additional questioning of teachers will lead to the school not having any teachers left. **Communities and parents thus feel that they struggle to hold teachers to account, since they do not have any bargaining power in the matter.** However, as with head teachers, the

---

[51] See further discussion on teacher motivation and morale in Chapter 8.

increased involvement of WECs and the government slogan of *"hapa kazi tu"* make parents and communities feel more confident asserting their opinions on education as they feel they have more to stand on.

Lastly, **in some case study sites the overall perception of parents and other respondents is that communities/parents do not have enough knowledge or understanding of schools' matters to be able to hold schools to account.** Instead both parents and teachers see it as schools' responsibility to guarantee quality of education and school development as they have the capabilities to do so, with one teacher stating: *"[the] community normally depends on [the] school to provide opinions…it is not possible for the opinions to come from the village government to school"* (school 2, district C). Some teachers believe that communities/parents do not assert influence in the school as *"they don't care about anything"* (school 2, district A). If in some cases this reflects the *status quo* of the communities' value of education, in other cases, it seems that school actors do not consider the opportunity costs linked to time and interest that parents have to engage with education.

References

Abadzi, H. (2006) Efficient learning for the poor: Insights from the frontier of cognitive neuroscience. World Bank, Directions in Development Series.

Bennell, P. and Akyeampong, K. (2007) Teacher Motivation in Sub-Sahara Africa and South Asia. Researching the Issues, 71, London, UK DFID.

Blundell, Richard, and Monica Costa Dias (2000) 'Evaluation Methods for Non-Experimental Data'. *Fiscal Studies* 21, no. 4: 427–68. http://www.jstor.org/stable/24437670.

Boyden J and Ennew J (eds) / Save the Children (1997) Children in Focus – a Manual for Participatory Research with Children. Stockholm: Save the Children Sweden

Caliendo, M. and Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. *IZA Discussion Papers*, No. 1588.

Cambridge Education (2014) Final EQUIP-Tanzania Inception Report, 5 February.

Cartwright, N, and Hardie, J., (2012) Evidence Based Policy: A Practical Guide To Doing It Better. Oxford: Oxford University Press.

Carvalho, S. and White, H. (1997) 'Combining The Quantitative and Qualitative Approaches to Poverty Measurement and Analysis', Technical Paper 366. Washington DC: World Bank.

Cueto, S et al. (2009) Psychometric characteristics of cognitive development and achievements instruments in Round 2 of Young Lives. Young Lives Technical Note #15, January 2009.

De Grauwe, A. (2001). School Supervision in Four African Countries: Vol. I: Challenges and Reforms. Paris: UNESCO IIEP.

De Ree, J. (2016). How Much Teachers Know and How Much it Matters in Class: Analyzing Three Rounds of Subject-Specific Test Score Data of Indonesian Students and Teachers. World Bank Policy Research Working Paper

Diseko, E., Barkhuizen, N., and Schutte, N. (2015) The Relationship between Talent Management and Turnover Intentions of Teachers in Botswana. 20th International Academic Conference

DFID (2013) Intervention Summary, Education Quality Improvement Programme in Tanzania (EQUIP-T). London, UK DFID.

DFID (2011) DFID Ethics Principles for Research and Evaluation

Dowd, A *et al.*, (2013) Literacy Boost Cross Country Analysis Results, Save the Children, Washington D.C.

EQUIP-T MA (2015) EQUIP-Tanzania Annual Report 2015. URT, PMO-RALG, MoEVT and DFID.

Garbarino, S., and Holland, J., (2009), 'Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results', Issues paper commissioned by DFID. GSDRC Emerging Issues Service.

Glewwe, P and Muralidharan, K. (2015) Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps and Policy Implications. RISE-WP-15/001. RISE Programme.

Glewwe, P, Kremer, M, and Moulin, S (2009). "Many children left behind? Textbooks and test scores in Kenya." American Economic Journal: Applied Economics 1 (1) (January 2009): 112-35

Guest, G., MacQueen, K., Namey, E., (2012). Applied Thematic Analysis. Sage Publications.

Hallinger, P., and Heck, R. (1996) The Principal's Role in School Effectiveness: An Assessment of Methodological Progress, 1980-1995. In K. Leithwood, *The International Handbook of Research on Educational Leadership and Administration*. New York: Kluwer Press.

Hardman, F, and Dachi, H. (2012) Evaluation of School-Based INSET Pilot Programme. York, UK: Institute for Effective Education, University of York.

Imbens, G. and Rubin, D. (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

Krippendorff, K., (2004). Content Analysis: An Introduction to Its Methodology. Sage Publications.

Mays, N., and Pope, C. (1995). Rigour and Qualitative Research, British Medical Journal 311(6997): 109-112.

Metzler, J and Woessmann, L. (2012). The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation. Journal of Development Economics. 99 (2)

MoEVT (2016) Curriculum for Basic Education Standards I and II. URT, MoEVT.

MoEVT (2009a) Basic standards for pre- and primary education in Tanzania. Dar es Salaam: MoEVT. www.ed-dpg.or.tz/pdf/PE/Basic%20Standards%20for%20Pre%20and%20Primary%20Education%20in%20TZ_2009.pdf

MoEVT (2005a) Mathematics Syllabus for Primary Schools Standard I-VII. URT, MoEVT.

MoEVT (2005b) Kiswahili Syllabus for Primary Schools Standard I-VII. URT, MoEVT.

Mulkeen, A. (2010) Teachers in Anglophone Africa. Issues in Teacher Supply, Training and Management. Washington, DC: The World Bank.

Nag S, Chiat S, Torgerson C, Snowling MJ (2014) Literacy, Foundation Learning and Assessment in Developing Countries: Final Report. Education Rigorous Literature Review. Department for International Development

OPM (2016a) EQUIP-Tanzania Impact Evaluation: Midline Planning Report. OPM.

OPM (2016b) EQUIP-Tanzania Impact Evaluation: Preliminary Indicators for the Programme Treatment Districts: Results from the Baseline and Midline Quantitative Surveys. OPM.

OPM (2015a) EQUIP-Tanzania Impact Evaluation Final Baseline Technical Report, Volume I: Results and Discussion. OPM. Available at: http://www.opml.co.uk/sites/default/files/OPM%20IE%20Final%20Baseline%20Report%20Volume%20I.pdf.

OPM (2015b) EQUIP-Tanzania Impact Evaluation Final Baseline Technical Report, Volume II: Methods and Technical Annexes. OPM. Available at: http://www.opml.co.uk/sites/default/files/OPM%20IE%20Final%20Baseline%20Report%20Volume%20II.pdf.

OPM (2016c) EQUIP-Tanzania Impact Evaluation: Propensity Score Analysis Review: A Technical Note. OPM.

OPM (2016d) EQUIP-Tanzania Impact Evaluation: Midline Quantitative Fieldwork Report. OPM.

Osim, R.O., Chika, C., Uchendu, I. O. (2012) Class size pressure: An Impediment to teacher's work. Global Advanced Research Journal of Educational Research and Review. Vol 1(5)

Ouane, A and Glanz, C (2011). Executive Summary. In Ouane, A and Glanz, C eds Optimizing Learning, Education and Publishing in Africa: the Language Factor. A Review and Analysis of Theory and Practice in Mother-Tongue and Bilingual Education in sub-Saharan Africa. UNESCO Institute for Lifelong Learning and the Association for the Development of Education in Africa / African Development Bank

Paternoster, R., Brame, R., Mazerolle, P. and Piquero, A. (1998), Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36: 859–866. doi:10.1111/j.1745-9125.1998.tb01268.

Rosenbaum, P.R. and Rubin, D.B. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician.* 1985: 39:33–38.

RTI (2016) Assistance to Basic Education All Children Reading (ABE ACR): Preliminary Findings Report, Tanzania National Early Grade Reading Assessment (EGRA). North Carolina. USAID.

RTI (2014) National Baseline Assessment for the 3Rs (Reading, Writing, and Arithmetic). Using EGRA, EGMA, and SSME in Tanzania. Study Report. Draft. Washington, DC: USAID.

Rubin, D. (2001) Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services & Outcomes Research Methodology* 2, 169–188.

Sabarwal, S., Evans, D., and Marshak, A. (2014). The permanent input hypothesis: the case of textbooks and (no) student learning in Sierra Leone. Policy Research working paper, no. WPS 7021. Washington, DC: World Bank Group

Siraj, I., Taggart, B., Melhuish, E., Sammons, P., and Sylva, K. (2014). *Exploring Effective Pedagogy in Primary Schools: Evidence from Research*. London: Pearson. Available at: https://research.pearson.com/content/plc/prkc/uk/open-ideas/en/articles/explore-eppse/_jcr_content/par/articledownloadcompo/file.res/Exploring%20Effective%20Pedagogy%20in%20Primary%20Schools.pdf

The National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research (1979) 'The Belmont Report': Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Department of Health, Education, and Welfare. [http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html]

The Open University (no date) Ethics Principles for Research Involving Human Participants, [http://www.open.ac.uk/research/main/sites/www.open.ac.uk.research.main/files/files/OU%20Ethics%20Principles%20for%20Research%20Involving%20Human%20Participant.pdf]

UNESCO (2016) If you don't understand how can you learn? GEM Report Policy Paper 24. Paris: UNESCO.

UNESCO (2015) Education For All Global Monitoring Report 2015: Education for All 2000-2015: Achievements and Challenges. Paris: UNESCO.

UNESCO (2014) Education For All Global Monitoring Report 2013/14: Teaching and Learning: Achieving Quality for All. Paris: UNESCO

US Department of Health and Human Services (2009) Code of Federal Regulations TITLE 45 Public Welfare. PART 46. Protection of human subjects. [http://www.hhs.gov/ohrp/policy/ohrpregulations.pdf]

Vogel, I., (2012), 'Review of the use of theory of change in international development', Review paper for the Department of International Development. Department for International Development.

Westbrook, J., Durrani, N., Brown, R., Orr, D., Pryor, J., Boddy, J., and Salvi, F. (2013) Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries. Final Report. Education Rigorous Literature Review. UK DFID. Available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/305154/Pedagogy-curriculum-teaching-practices-education.pdf

White, H. (2009) Theory-Based Impact Evaluation: Principles and Practice. 3ie Working Paper 3. International Initiative on Impact Evaluation.

Wright, B and Stone, M. (1979) *Best Test Design*. Chicago: Mesa Press.

ANNEXES

Annex A   IE districts

Annex table 1:   Impact evaluation districts

Control/treatment	Region	District
Control regions and districts in IE study	Arusha	Ngorongoro DC
	Mwanza	Misungwi DC
	Pwani	Rufiji DC
	Rukwa	Nkasi DC
	Ruvuma	Tunduru DC
	Singida	Ikungi DC
		Singida DC
	Tanga	Kilindi DC
Treatment districts in IE study (Note: all 17 districts are part of the quantitative survey, * indicates they are also part of the qualitative research)	Dodoma	Bahi DC
		Chamwino DC
		Kongwa DC
		Mpwapwa DC *
	Kigoma	Kakonko DC
		Kibondo DC
	Shinyanga	Kishapu DC *
		Shinyanga DC
	Simiyu	Bariadi DC
		Bariadi TC
		Itilima DC
		Maswa DC
		Meatu DC
	Tabora	Igunga DC
		Nzega DC
		Sikonge DC
		Uyui DC *
Treatment districts that are not part of the IE study (Note: districts in Lindi and Mara joined EQUIP-T in 2015)	Dodoma	Chemba DC
		Kondoa DC
	Kigoma	Buhigwe DC
		Kasulu DC
		Kigoma DC
		Uvinza DC
	Shinyanga	Kahama DC
		Msalala DC
		Ushetu DC
	Simiyu	Busega DC
	Tabora	Kaliua DC
		Uramba DC
	Lindi	Kilwa DC
		Lindi DC
		Liwale DC

		Ruangwa DC	
		Bunda DC	
		Butiama DC	144
	Mara	Musuma DC	
		Musuma MC	
		Rorya DC	

Annex B   Stakeholder engagement and IE governance

B.1   Stakeholder engagement

Stakeholder engagement is an ongoing process as part of the IE, and started from the inception phase with the overall design of the IE. Volume II of the BL IE report sets out the stakeholder consultations carried out in inception and in disseminating the BL findings. Plans for engagement and dissemination were set out and agreed in the IE ML Planning Report (OPM 2015a).

For the ML, stakeholder engagement began in October 2015 with an application to Tanzania's Commission for Science and Technology (COSTECH) to approve the ML research. The update to the IE design for the ML round started in November 2015 with a small IE team (comprising OPM staff and enumerators from the BL round) visiting two schools to explore the effect at school level of some of the changes in education policy that had taken place since BL. This was followed by a visit in January 2016 by the lead researchers from IE team to attend the EQUIP-T programme mid-term review and to conduct follow-up interviews with EQUIP-T, LANES and government officials. This preparatory engagement was used to understand the status of EQUIP-T (and LANES) implementation, and to gather expectations of changes that might already be seen, in order to adjust aspects of the IE design for ML. Since then, engagement has been regular with DFID to update on the IE team's progress, with the study's Reference Group (including MoEST and PO-RALG, see section below on this) to share early findings for comments, and with the MA to facilitate data collection and avoid over-burdening interviewees. The main consultations and missions carried out since BL are shown in Annex table 2.

Annex table 2: Stakeholder consultations and visits

Date	Purpose
October 2015	• Application made to by COSTECH for the IE (subsequently granted).
November 2015	• School visits to explore the effect of changes in education policy (new curriculum and changes to capitation grant mechanism). • Testing of quantitative instruments in updated software (pre-test 1).
January 2016	• Attended the EQUIP-T mid-term review steering committee meeting. • Taking note of the discussions between EQUIP-T and a wide range of sector stakeholders including senior government officials and representatives from other education programmes • Meet with DFID Education Advisor and all Component Leads at the EQUIP-T MA to understand implementation progress, plans for the coming months, and expectations of changes so far. • Meet with the GPE LANES coordinator to receive information on GPE LANES implementation progress and implications for EQUIP-T and control districts.
February 2016	• Meetings with the Commissioner for Education, Professor Eustella Bhalalusesa, and with the Director of Primary Education, Ms Sarah Mlaki, to consult them on the ML research focus, and to review and update the IE Reference Group membership. • Pre-test 2 using CAPI survey instruments by the impact evaluation team. • Visits to NECTA and TIE to brief them on the ML IE, and to invite representatives to join the IE Reference Group.
March/April/May 2016	• ML IE Planning Report submitted to the Reference Group for comment and feedback. • Supervisor and enumerator training and pilot for quantitative BL survey by the impact evaluation team. • Quantitative data collection.
April/May 2016	• Briefing given to the incoming EQUIP-T National Programme Coordinator. • Qualitative research training. • Qualitative data collection with the EQUIP-T MA, PO-RALG, regions, districts and WECs. • Qualitative data collection with nine case study schools.

June 2016	• Meetings with EQUIP-T MA for data collection for the costing study. • Visit by project manager to supervise the manual marking of the TDNA.
July 2016	• ML IE Preliminary Quantitative Descriptive Trends Report submitted to the Reference Group for comment and feedback.
September 2016	• Draft ML IE Preliminary Issues Note on Teacher In-Service shared with DFID for comment
October/November 2016	• Telephone briefing with incoming DFID Education Advisor. • Phone call with new Acting Commissioner for Education, Mr Nicholas Buretta.
November 2016	• Sharing IE ML findings at the Joint Education Sector Review.
December 2016	• Presentation of IE ML findings at the EQUIP-T Annual Review steering committee meeting. • Presentation of IE ML findings at the IE Reference Group. • Presentation of IE ML findings at GoT/Development partners meeting on harmonising approaches to teacher INSET.
2017 (planned)	• Paper accepted for African Evaluation Association Conference (8th AfrEA International Conference, Uganda). • To be confirmed. The LANES literacy and numeracy conference managed by USDM in 2015 provided a good forum for dissemination of BL findings. This may be held again in 2017. The UKFIET Oxford Conference on Education and Development 2017 is another good forum to share findings.

The principal audience for this midline IE study are EQUIP-T's Managing Agent, DFID, and GoT officials involved with the programme. It is intended that the evidence will inform discussions on programme adjustment at this half-way point in implementation. The study will also serve a wider range of education sector stakeholders who are interested in recent changes in primary education is some of the most deprived areas of Tanzania. Plans for dissemination to these stakeholders were set out in the ML Planning Report (OPM 2015a) and include sharing the findings in this report at the Joint Education Sector Review (November 2016), the EQUIP-T Annual Review steering committee (December 2016) and the IE Reference Group (December 2016).

All of the reports, briefing notes, issues papers and other products produced as part of the IE are available on OPM's website (www.opml.co.uk), (see for example: http://www.opml.co.uk/sites/default/files/OPM%20IE%20Final%20Baseline%20Report%20Volume%20I.pdf.). The BL briefing notes and conference papers produced using the IE findings have all been uploaded on to the Social Science Research Network www.ssrn.com (see for example: https://ssrn.com/abstract=2779240; https://ssrn.com/abstract=2579284; and https://ssrn.com/abstract=2782747). ML products will also be uploaded to these websites.

The BL quantitative survey data (anonymised) is publically available on the World Bank microdata library http://microdata.worldbank.org/index.php/catalog/2290 The ML quantitative data will also be documented and uploaded on to this catalogue. The IE team is in contact with a number of university researchers who are interested in using the data for further research and publication.

B.2   Reference Group (RG)

At the start of the IE in 2014, the Ministry of Education led a process to form an EQUIP-T IE RG to provide technical recommendations and feedback to the OPM IE team. The terms of reference for the RG are included in the Midline Planning Report (OPM 2015a, Annex F). At BL, the RG held its first meeting to review and comment on the overall IE design (January 2014). A second RG meeting was held in November 2014 where baseline findings were discussed extensively, feedback provided to guide revisions to the report, and members advised the IE team on opportunities for dissemination as well as links with other studies and programmes. The RG will continue to perform this oversight and guidance role for the midline IE. The RG is requested to provide feedback on all deliverables, and so far two preliminary reports from the ML research have been shared for comment (see Annex table 2). A third RG meeting took place in December 2016 to

discuss the draft midline IE report and plan for dissemination of the findings. During the meeting, members provided useful feedback on the draft report (noted in the meeting minutes, subsequently circulated to all members for corrections or additions), and members were also requested to provide any additional feedback in writing. The evaluation team consolidated all the feedback received on the draft report from DFID, the EQUIP-T MA, and other RG members into a document. From this, the team carefully considered each comment and made changes to the draft report where appropriate. The team also drafted a written response to each comment, explaining how the comment had been dealt with in the final report or justifying why no changes had been made. This commentary was submitted to DFID together with the final draft report.

From its inception until mid-2016, the then Commissioner for Education, Professor Bhalalusesa, chaired the RG.[52] Members of the IE team met with the Commissioner in February 2016 to discuss the midline IE, and a proposal from the IE team to slightly alter the composition of the RG to better represent the range of stakeholders active in primary education in Tanzania. New members accepted the invitation to join the RG. The RG is convened by Professor Herme Mosha from UDSM who is a core member of the IE team. The organisations represented on the IE RG are:

- MoEST (Commissioner of Education; Director of Primary Education; District Education Officer)
- PO-RALG (Director of Primary Education)
- National Examinations Council of Tanzania (Senior Examinations Officer)
- Tanzania Institute of Education (Director for Training)
- DFID (Education Advisor; Results Advisor)
- EQUIP-T Managing Agent (National Programme Coordinator)
- LANES (National Programme Coordinator)
- USDM, School of Education (Associate Dean/Senior Lecturer; Lecturer/Coordinator Graduate Studies)
- Twaweza East Africa (Senior Consultant; Advisor)

B.3 Impact evaluation governance and quality assurance

Oversight and policy direction for the impact evaluation is provided by an OPM Governance Team comprising the OPM Managing Director, the OPM Director of Statistics, Evidence and Accountability, the OPM education portfolio lead, and an OPM Education Associate who is Senior Education Advisor in the IE core senior team (see below).

Management is executed by the Project Manager, an OPM Principal Education Consultant, who in addition to playing a leading technical role is responsible for team management, the coordination of inputs, financial management and liaison with the supporting administration team and research teams in OPM's Oxford office and OPM's Tanzania Office respectively, and OPM's internal reporting and project oversight processes.

The Project Manager is responsible to the OPM Governance Team for successful delivery of the impact evaluation. The Project Manager is supported by a core senior team and a wider team of technical specialists (see Annex table 3 below). The core senior team comprises, a deputy project manager, a senior education advisor (also part of the OPM governance team), and a senior national education advisor. There are 14 technical specialists in the wider technical team. The project manager ensures that the two teams work together to meet the objectives of the evaluation,

---

[52] It is hoped that the new Acting Commissioner, Nicholas Buretta, will agree to take over this role.

and to produce the key deliverables. The core team has is responsible for stakeholder engagement including dissemination of findings and engagement with the RG.

Annex table 3 ML IE team members and roles

Name	Role
Georgina Rawle	Project Manager/ML design /Pupil learning assessment Lead
Nicola Ruddle	Deputy Project Manager/ ML design /Costing Study Lead
Paud Murphy	Senior Education Advisor
Professor Herme Mosha	Senior National Education Advisor
Dr Gunilla Pettersson Gelander	Senior Education Specialist/Descriptive Quantitative Analysis Lead
Andreas Kutka	Quantitative Survey Fieldwork Lead
Ignatus Jacob	Quantitative Survey Fieldwork Manager
Diego Shirima	Quantitative Survey Data Manager
Jana Harb	Quantitative Survey Fieldwork Deputy & Data Analyst
Stephi Springham	Quantitative Data Management & Documentation Specialist
Jean Davis	Quantitative Survey Analyst
Michele Binci	Impact Estimation Lead/ML design
Paul Jasper	Senior Impact Estimation Analyst
Madhumitha Hebbar	Impact Estimation Analyst
Shrochis Karki	Qualitative Research Advisor/ML design
Johanna Wallin	Senior Qualitative Researcher
Deogardius Medardi	Qualitative Fieldwork Manager/Quant Fieldwork Team Member
Alice Aldinucci	Qualitative Research Analyst

Quality Assurance for the ML research is provided using a number of layers of review. In the first stage, each key activity and output has been reviewed internally by other project team members, led by Georgina Rawle and Gunilla Pettersson Gelander for quantitative descriptive analysis, Michele Binci and Paul Jasper for impact analysis, Nicola Ruddle and Johanna Wallin for qualitative analysis, and Nicola Ruddle and Georgina Rawle for the costing study. The qualitative analysis was further reviewed by OPM qualitative experts Dr Shrochis Karki and Dr Marlene Buchy. The learning outcomes analysis including Rasch modelling and construction of interval scales was reviewed by Dr Joshua McGrane (Psychometrician and Rasch measurement specialist, University of Western Australia).

In the second stage, the full drafts of Volume I and Volume II were shared with three reviewers: Paud Murphy, Senior Education Advisor, Professor Herme Mosha (University of Dar es Salaam), Senior National Education Advisor, and Dr Caine Rolleston a leading academic researcher in the field of education and economics (Institute of Education, University College London).

A final stage of external quality assurance is provided through the Impact Evaluation Reference Group meetings, review and feedback by SEQAS and DFID as well as further external review by technical experts of each of the IE key outputs, as appropriate.

Annex C   Ethical considerations

C.1   Principles

As this research involves human subjects, it is important to be fully aware of the ethical considerations. A review of best practice was conducted to inform the design and protocols of the midline fieldwork and data use. This review looked at the protocols OPM used in the baseline, those used in OPM's other education evaluations, those used by other research organisation in Tanzania, and guidance from organisations specialising in children's rights (Save the Children, 2007), research (Open University, US Department of Health and Human Services) and development (DFID, 2011).

There are three basic ethical principles of research with human subjects, as set out in the Belmont Report (1979):

1.   Respect for persons
2.   Beneficence
3.   Justice

The section below outlines how the midline research adheres to each of these principles.

C.1.1 Respect for persons

This means the prospective participants should be given the information they need to decide whether or not they want to participate, they should be given the freedom to decide not to participate or to stop at any point. In particular, this means that participants should give informed consent, agreeing to take part voluntarily and with adequate information. Where a participant has diminished autonomy – in this case children – they are entitled to additional protections.

In both the quantitative and qualitative surveys, all participants were read a statement before the interview/focus group began. The statement sets out what the research is for, how and why they were selected for the discussion, the confidentiality of their responses, how responses will be used (and in particular that they will not affect their grades or job), that the process is optional and they are welcome to ask questions or leave at any time. After this, the enumerator/interviewer/ facilitator asked them if they agree to continue. At this point, the enumerator ticks a box (in CAPI or hand-written for qualitative sessions) to confirm the participant has given oral informed consent to continue. In the qualitative discussions, participants had the right to refuse audio recording.

Where children are being interviewed, the head teacher was asked to give consent on behalf of the parents, and children were also given the opportunity to decide not to take part (assent). This consent statement and agreement was done collectively for the children involved in the qualitative fieldwork, and individually for the quantitative fieldwork, away from other teachers or parents, so that they do not feel pressured either way. The head teacher also gave consent for researchers to observe a class, as did the teacher of that class, and the teacher asked pupils if they were ok with researchers sitting in on their class.

Oral consent was chosen based on OPM's experience with research in developing countries, and in particular with respondents who are not literate and/or are not familiar with research. In these cases, respondents become very formal and often even worried if asked to sign a piece of paper. Respondents need to be as relaxed and responsive as possible, so allowing them to consent orally

and recording it meticulously (by the research team) achieved the same function without compromising the quality of our interactions.

C.1.2 Beneficence

This principle requires that no harm is caused by the research. There are a number of ways in which the research team adhered to this principle. Participants were interviewed in an environment which the participants are comfortable with, and secures them privacy. Responses are confidential – respondents will not be named and the specifics of who was interviewed or who gave specific responses was not given to anyone outside of the research team. This means that no responses are attributable, and the report does not include responses that are traceable. These principles are intended to avoid any social risk from views being overhead by others in the community or those above them in the reporting line, and should allow respondents to speak more honestly. It was explained to participants that there would be no personal repercussions from taking part or what they say. The quantitative data set will be made publically available but anonymised. The research teams were trained in confidentiality. Finally, the discussions were organised to minimise cost or disruption for the participants.

Particular care is taken given the engagement with children. The research involves interviewing children in standard 3, who generally are between the ages of 9 and 11 years. Given their age, it is important they are treated with care and respect, and given full opportunity to decide to opt out of the work. The fieldworkers carrying out the interviews (both quant and qual) were trained on the ethics of working with children – ensuring a safe and private space for their participation, letting them ask questions, making it clear it is fine for them to leave a question or leave the interview entirely, keeping responses confidential and anonymous – verbally but also by carefully handling the data collected. These processes were set out in the enumerator manuals used during training and be available for reference during the fieldwork. No responses should be coerced, participants are free to not respond.

C.1.3 Justice

Justice requires that individuals and groups are treated fairly and equitably. In this case, there is no notable benefit (except refreshments in a focus group discussion) or burden (except time) of taking part in the research, and all participants will be subject to the same benefits and burdens. In the qualitative fieldwork in particular, participants are selected in a way that there is no exclusion based on biases and participants must understand how they have been selected.

C.2 Process

The OPM Ethical Review Committee approved the fieldwork proposals, including the instruments and the informed consent statements.

OPM have gained approval for this research from the Tanzania Commission for Science and Technology (COSTECH) which has the mandate of co-ordinating and promoting research and technology development activities in the country.

C.3 Conflict of interest considerations

The OECD DAC standards related to potential conflict of interest arising in development evaluations are listed in the table below, together with a commentary on their relevance to this EQUIP-T IE.

Annex table 4 Potential conflict of interest considerations and response

Quality standards for development evaluations related to conflict of interest [1]	Commentary on relevance to the EQUIP-IE process
Evaluators are independent from development interventions including its policy, operations and management functions, as well as intended beneficiaries.	**Independence from EQUIP-T policy, operations and management** The IE team has not been involved in the policy, planning or implementation of the EQUIP-T programme. The IE team made its judgements and drew conclusions from the evidence collected, and was not influenced or pressured by external parties. The IE team's interactions with the EQUIP-T MA were limited to discussions and requests for information to inform the evaluation, but the IE team chose how to use the information freely. **Independence from EQUIP-T beneficiaries** EQUIP-T's target beneficiaries are pupils, teachers and education administrators. The IE team's interactions with these parties was limited to data collection, and the representation of findings derived from this evidence based on objective research criteria.
The evaluation team is able to work freely without interference. It is assured of co-operation and access to all relevant information.	The IE team did not experience any barriers to working freely and without interference. Requests for information from the EQUIP-T MA have been met to the extent possible from their emergent information systems. Response rates were generally very high in the fieldwork.
Possible conflicts of interest are addressed openly and honestly.	OPM holds the contract for the EQUIP-T IE, and at the same time OPM is subcontracted to the EQUIP-T MA contract-holder (Cambridge Education) in other programmes in other countries. In principle this could pose a conflict of interest but in reality this is not the case. OPM is fully committed to, and has a long track record of, conducting high-quality independent research. Part of OPM's operating philosophy is 'independence in our judgements' (see http://www.opml.co.uk/about-us/our-values). As is standard with OPM's approach, the work on this IE is subject to peer review from highly regarded academics (see Annex B.3) to ensure the technical quality of the work and the validity of the recommendations given the evidence.

Notes (1) Source is OECD DAC (2010, p11).

Annex D   Quantitative data collection

OPM's Tanzania office conducted the ML IE survey. A detailed report on the fieldwork is available (OPM 2016d), including a section on lessons learned which will serve as a starting point for planning the endline survey in 2018. This annex summarises the key points from the fieldwork report.

D.1   Personnel

The fieldwork management team comprised seven members (including six OPM staff) led by a quantitative survey project manager who had overall responsibility for the design, implementation, management and quality of the fieldwork. Since almost all the survey instruments were administered using computer assisted personal interviewing (CAPI), the team also included several members with very strong computer programming skills in the relevant software (Surveybe). The overall project manager for the IE, who is responsible for the content of the instruments worked closely with the fieldwork team during pre-testing, training, piloting and early fieldwork.

 51 enumerators were invited to the training. These were selected based on the following criteria (in order): (i) good performance during the EQUIP-T BL survey (about half of the enumerators from BL confirmed for ML); (ii) interviewers with strong track record from other OPM-led surveys; (iii) new recruits–these were interviewed over 2 days and selected based on their prior survey experience and knowledge of education.

D.2   Fieldwork preparation

The early fieldwork preparation consisted of pre-testing the instruments and protocols, obtaining permits from the government for visiting schools during the pre-tests, pilot and fieldwork, revising the BL fieldwork manual, and refining the instruments and protocols.

D.2.1   Pre-test 1

The first ML pre-test took place in Kinondoni district (Dar es Salaam) on November 24th-25th 2015. A small team of two OPM staff members and three enumerators from the BL survey visited 2 schools to: i) test the functionality of the updated of the electronic questionnaires in the updated CAPI software (Surveybe); ii) gather information on how the change in government, the introduction of the Literacy and Numeracy Educational Support (LANES) Programme in 2015 and the resulting change in Standard 1 and 2 curriculum were affecting primary education at school level.

D.2.2 Pre-test 2

A full pre-test of all instruments and protocols took place from the 8th to 12th of February 2016 in Kisarawe District, Pwani region. A team of 15 (five OPM staff, one OPM intern, seven enumerators, a DFID representative, and an education professor from the University of Dar es Salaam who is a senior member of the IE team) visited four schools, following two days of classroom based training. The pre-test resulted in the following outcomes:

- Refinement of the instruments and data collection protocols;
- Refinement of the translation of instruments from English to Kiswahili;
- Significant changes made to the development of the instruments in CAPI (Surveybe);

- Development of protocols for tracking households and ethical protocols for escorting pupils at the end of the day to their home; and
- Decisions made on instruments to be conducted in control vs. treatment schools, training timeline and broad plan, fieldwork timeline, fieldwork model, and team composition.

D.2.3 Permits and reporting

As part of preliminary preparations for any survey in Tanzania, there are two types of governmental permits that have to be obtained prior the beginning of Research work:

- **COSTECH Permit** - Mandatory for any research activity in Tanzania.
- **Ministry Permit** - Different partners in the field require Ministry letters, as few recognise COSTECH. These permits gives the order to local administration to cooperate with the research and support the field teams.

Upon receipt of the permits, the anticipated field work needs to be reported at the regional and district level. Letters introducing the study to local leaders are obtained in the process.

For the ML IE survey, the COSTECH research clearance and an introduction letter was received two months prior the start of actual fieldwork.

For the Ministry permits, OPM reported to The Prime Minister's Office Regional Administration and Local government (PMORALG) and to the Ministry of Education and Vocational Training (MoEVT). Reporting to MoEVT was relatively fast and simple. The initial submitted letters were followed up in person, and an introduction letter to all 12 Regional Administrative Secretaries was received after 7 days. Getting government approvals from PMORALG proved to be challenging and very time-consuming. The various stages are laid out in the full fieldwork report (OPM 2016d), but the final decision was to shift to physical reporting approach, as sending letters by courier and follow-up phone calls were unsuccessful. In a combined effort, three of the fieldwork management team reported in person to all 10 regional and 25 district offices during the enumerator training period. In total 50 person days (including travel days, as distances are vast) had to be allocated to this final reporting task.

D.2.4 Fieldwork manual

Using the BL fieldwork manual as a basis, an extensive field worker manual was developed that covered basic guidelines on behaviour and attitude, the use of CAPI and data validation procedures, instructions on fieldwork plans and procedures (sample, targets, replacements, communication, and reporting) as well as a dedicated part on the description of all instruments and protocols. Insights from the pre-test were reflected in the manual.

Draft versions of the instrument and protocol sections of the manuals were printed, handed out to interviewers as a reference during the training, and used as guidelines by the trainers. The manual was updated on an ongoing basis during the training and pilot phase where updated conventions or additional clarifications were needed. The final version of the manual was printed at the end of the pilot phase and copies provided to the field teams.

D.3   Training and pilot

Enumerator training and a field pilot took place in Dar es Salaam and Dodoma from 29th March to 14th April. A total of 47 enumerator trainees participated in the training. The training was delivered by four members of the fieldwork management team and the overall IE project manager.

The main objective of the training was to ensure that team members would be able to master the instruments, understand and correctly implement the fieldwork protocols, comfortably use CAPI, and be able to perform data validation. Supervisors were furthermore trained on their extra responsibilities of data management, fieldwork and financial management, logistical tasks, and the transmission of data files to the data manager.

The training had two components: a classroom-based training component and a field-based component that included a full scale pilot. The performance of enumerators was assessed on an on-going basis, using written assessments and observation of performance in the field and these scores were recorded. At the end of the training and pilot phase, the final fieldwork team was selected using this information.

D.4 Fieldwork organisation

D.4.1 Fieldwork plan

The fieldwork plan was designed to cover all 200 schools within all 12 regions and 25 districts for the duration of not more than 7 weeks starting April 15th 2016 to May 27th 2016. Teams communicated regularly with OPM to report delays and/or any event likely to affect the feasibility of the fieldwork plan.

D.4.2 Fieldwork model

The team composition and field work model at ML were set up differently to BL to: a) reduce transport costs by reducing car days relative to field worker days and moving more travel days to Saturday (schools closed, but working day for field workers), and b) to be able to translate the reduced requirements of instruments in control schools into reduced team size for control teams. At BL, fieldwork was undertaken by 15 teams of 3 field workers each visiting a school on two consecutive days. At ML 4 treatment teams of 6 field workers and 4 control teams of 5 enumerators visited one school on one day.

D.5 Fieldwork implementation

The fieldwork started on the 15th of April and ended on the 27th of May with no major breaks in-between.

D.5.1 Replacements

D.5.1.1 Schools

All schools that were interviewed at BL were revisited and interviewed at ML, and hence no replacement of schools took place. There were only 3 cases where enumerators visited a school and were unable to conduct the survey because of an event that was taking place at the school on the day. These events included a S7 mock exam, BRN exams and a vaccination day. In that case, the teams rearranged to come back another day to conduct the survey in those schools.

D.5.1.2 Pupils and teachers

Only 64 pupils (out of 2,972 pupils) were replaced. The reasons are: 2 refused, 42 were unavailable due to sudden events such as illness, 18 were absent (but had been recorded by the teacher as present and hence were part of the sampling frame), and 2 were mistakenly recorded in the Standard 3 register when they were actually in Standard 2.

No replacement was done for the teacher interviews, as no sampling was required. 6 teachers were replaced for the Kiswahili TDNA, 4 teachers were replaced for the Maths 1-3 TDNA, and 9

teachers were replaced for the Maths 4-7 TDNA. The most common reason for replacement was absenteeism.

D.5.2 Response rates per instrument

Table 2 in Chapter 3 shows the generally high response rates for each instrument. Here is some further information underlying the response rates for selected instruments:

- If the parent of the tested pupil could not be reached, as a last resort, the poverty score card could be administered to the pupil. This happened in 216 out of 2,963 cases (7%). Some of the reasons given by enumerators were that the pupil is boarding and parents live far away, pupil lives too far away to be reached, and parents were not found at home.
- Some 70 of the 817 teacher interviews (9%) were conducted over the phone, as the teacher was absent on the day of the survey.
- In 36 out of the 200 schools, the HT or Acting HT was absent on the day of the survey and as a result another teacher was interviewed instead to collect information related to school records. After fieldwork ended, head teachers in 26 of those 36 schools were reached over the phone for interview.

D.6 Quality control and data checking protocols

At the end of each working day, supervisors collected all interview files from their team members and uploaded them into a shared and organised Dropbox folder that was set up by the data manager. The data manager would receive all files from all 8 teams and export them into Stata data files (a statistical programme) and then run daily checks on all files to make sure they are complete and identify potential errors.

Several mechanisms were put in place in order to ensure high quality of the data collected during the survey. These are briefly summarised in turn below:

D.6.1 Selection and supervision of enumerators

As discussed above, each enumerator was supervised at least once by the training team during the training, piloting and first week of data collection. This allowed a well-informed selection of enumerators and their allocation into roles matching individual strengths and weaknesses.

D.6.2 CAPI built-in routing and validations

One important quality control means in CAPI surveys are the use of automatic routing and checking rules built into the CAPI questionnaires that flag simple errors during the interview, i.e. early enough for them be corrected during the interview. In each CAPI instrument, validations and checks were incorporated in the design in order to significantly reduce errors and inaccuracies during data collection. In addition to having automatic skip patterns built into the design to eliminate errors resulting from wrong skips, the CAPI validations also checked for missing fields, out of range values and inconsistencies within instruments.

D.6.3 Secondary consistency checks and cleaning in Stata

The ML survey exploited another key advantage of CAPI surveys, the immediate availability of data, by running a range of secondary consistencies checks across all data on a daily basis in Stata. Data received from the field was exported to Stata the following day, and a range of do-files

were run to assess consistency and completeness, and make corrections if necessary. The checks comprised the following:

ID uniqueness and matching across instruments

Completeness of observations: target sample size versus actual

Intra and inter-instrument consistency and out of range checks

The data manager ran the checking do-file on a daily basis on the latest cleaned data. This would return a list of potential issues in the long format which the data manager would then investigate and undertake the necessary cleaning actions. Whenever any issue was flagged, effort to obtain an explanation was undertaken either by reviewing enumerator comments or phoning teams.

In addition to the checking and cleaning process, all enumerator comments as well as other specify variables were translated from Swahili to English. All translated entries were further reviewed by the data analysis team in order to 1) ensure that they are understandable and properly translated into English and 2) none of the other specify answers for multiple response questions are in fact synonymous to one of the response items. The revision resulted in a long list of other specify items that were then recoded into one the available response items.

D.6.4 Monitoring fieldwork progress and performance indicators

In addition to the above checks that were specific to each instrument, the survey team monitored the general progress of the fieldwork and specific indicators revealing the performance of teams and enumerators over time. Indicators such as number of control/treatment schools completed, number of teachers/pupils/parent/lesson observations interviews completed, average interviewing time of each instrument, number pupils were interviewed instead of their parents, how many teacher interviews were conducted over the phone, etc. These indicators were constructed in a Stata do-file that ran on the latest cleaned dataset and was then uploaded onto a google document sheet that would break down each of the indicators by team, enumerator (where applicable) and week of data collection. This was reviewed regularly by the fieldwork management team, and overall IE project manager, and used to feedback to weaker teams and to improve performance.

D.6.5 Back-checking data

The QA protocol involved back-checks that were conducted over the phone and in the field. Two members of the fieldwork management team called back interviewed teachers to confirm that the interviews were indeed conducted. Furthermore, a list of questions to be re-asked to teachers were compiled and administered to the teacher over the phone to ensure that the information was properly collected. In addition, the fieldwork management team re-visited 10 schools and 45 households to check whether interviews were administered properly.

D.6.6 Integration of Analysis and Survey Team

Another central element of quality assurance was the strong integration of the fieldwork management team and the members of the quantitative analysis team, including the overall IE project manager. Member of both teams were involved in the fieldwork preparation and implementation, and in the analysis process which followed.

D.7   Fieldwork challenges and lessons learned

The EQUIP-T ML field work experienced several challenges. Some of the key lessons to be considered for the next round of data collection in 2018 are:

- Assess the (analytical) feasibility of changing the field work window to early February and completion of field work prior to the midterm breaks. Collect the region specific term times from the REO as early for planning purposes.

- Plan physical regional and district level reporting from the beginning and ideally complete it prior to the training. A dedicated person should lead the reporting to avoid overstretching survey management. Timely COSTECH and ministry clearances are necessary.

- Follow the protocol used at ML to avoid visiting schools when they are closed or holding events. The ML teams contacted the DEOs and Head Teachers one week in advance (to explore and confirm school timetables and accessibility "in the next month"). The teams did not disclose precisely the date of the visit to avoid schools "preparing" for the visit and to not distort the data collected on teacher attendance and punctuality.

- Follow the approach used at ML to mitigate the impact of the rain risk. For example, supervisors were instructed to check for potential access issues to schools the day prior to visiting.

- To minimise refusals in the TDNA exercise, present it to teachers as a marking exercise rather than a "test".

- Take into account that in certain areas, particularly Ngorongoro and Simiyu, there are many non-Swahili speakers. This can make it difficult to interview parents. Some teams at ML used teachers who knew the vernacular language to translate for the scorecard interview with parents.

- Take into account the poor internet and phone coverage in some areas (especially Ngorongoro district at ML). Provide routers with different networks to try to mitigate risk.

- Train all enumerators on testing pupils, to maximise flexibility in teams.

- Have a slightly larger pool of trainees, so that additional trained enumerators are available in case of attrition.

- Add a training facilitator to the fieldwork management team to support the training logistics.

Annex E   Definitions of key quantitative indicators

E.1   Chapter 2 Pupil Learning and background characteristics

Indicator name	Indicator definition	Respondent / unit of analysis	Notes
Pupil learning in Kiswahili			
Stds 3 pupils in Kiswahili performance band x (%)	Number of Std 3 pupils with ability scores that fall in or on the boundary of Kiswahili performance band x/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	Estimates of pupil ability and item difficulty are estimated using Rasch analysis (item-response theory modelling). Both are mapped on to a common scale. The items relate to statements in the standard one and standard two curriculum, and can be used to draw performance band boundaries to mark, for example, the increasingly difficult skills required to move from one curriculum level to another. The performance band boundaries are defined using estimates of item difficulties linked to curriculum competencies and mapped on to the same scale as the pupil ability estimates.
Std 3 pupil Kiswahili test score (% correct answers)	Number of Kiswahili questions answered correctly/total number of questions, expressed as a percentage, for each Std 3 pupil.	Standard 3 pupils	These indicators are constructed using raw-score data. Non-response is treated as incorrect in the scores.
Pupil learning in mathematics			
Stds 3 pupils in mathematics performance band x (%)	Number of Std 3 pupils with ability scores that fall in or on the boundary of mathematics performance band x/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	Estimates of pupil ability and item difficulty are estimated using Rasch analysis (item-response theory modelling). Both are mapped on to a common scale. The items relate to statements in the standard one and standard two curriculum, and can be used to draw performance band boundaries to mark, for example, the increasingly difficult skills required to move from one curriculum level to another. The performance band boundaries are defined using estimates of item difficulties linked to curriculum competencies and mapped on to the same scale as the pupil ability estimates.
Std 3 pupil mathematics test score (% correct answers)	Number of mathematics questions answered correctly/total number of questions, expressed as a percentage, for each Std 3 pupil.	Standard 3 pupils	These indicators are constructed using raw-score data. Non-response is treated as incorrect in the scores.
Pupil background characteristics			
Main language spoken at home not Kiswahili (% Std 3 pupils):	Number of Std 3 pupils reporting that the main language spoken at home is not Kiswahili/all Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Pupils below poverty line (% Std 3 pupils)	Number of Std 3 pupils that come from a poor household/all Std 3 pupils, expressed as a percentage.	Standard 3 pupils	A pupil is considered 'poor' if he/she comes from a household that has a greater than 50% probability of being below the Tanzania national poverty line, and 'rich' otherwise.
Ate before school (% Std 3 pupils)	Number of pupils reporting that they ate something before school on the day of the survey/all pupils, expressed as a percentage.	Standard 3 pupils	
Books, newspapers at home (%)	Number of Std 3 pupils reporting that there are books, newspapers or other reading materials available in their home/all Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Time to school (mean minutes)	Average time to get to school in minutes for Std 3 pupils.	Standard 3 pupils	

| Mother cannot read & write (%) | Number of Std 3 pupils' mothers that cannot read and write/all Std 3 pupils, expressed as a percentage. | Standard 3 pupils | |

E.2 Chapter 3 Teacher capacity, performance and conditions for learning

Indicator name	Indicator definition	Respondent/unit of analysis	Notes
Chapter 3 EQUIP-T component 1: Teacher capacity, performance and conditions for pupil learning			
Provision of teacher INSET (EQUIP-T input)			
Attended EQUIP-T in-service training last two years (% Stds 1-2 teachers)	Number of teachers of Stds 1-2 that report attending EQUIP-T in-service training the previous two years/all interviewed teachers of Stds 1-2, expressed as a percentage.	Standards 1-2 teachers	The relevant period for BL is 2012-2013 and for ML 2014-15.
Has teacher INSET coordinator (% schools)	The number of head teachers responding that the school has a coordinator for teacher in-service training/all interviewed head teachers, expressed as a percentage.	Head teachers	
Received EQUIP-T in-service training away from school (% Stds 4-7 teachers):	Number of teachers of Stds 4-7 that received EQUIP-T in-service training away from school the previous two years as reported by head teacher/all teachers of Stds 4-7 currently working at the school regardless of standards taught, expressed as a percentage.	Standards 4-7 teachers	
Received school-based EQUIP-T in-service training (% Stds 4-7 teachers):	Number of teachers of Stds 4-7 that received school-based EQUIP-T in-service training the previous two years as reported by head teacher/all teachers of Stds 4-7 currently working at the school regardless of standards taught, expressed as a percentage.	Standards 4-7 teachers	
Received EQUIP-T in-service training away from school (% Stds 1-2 teachers):	Number of teachers of Stds 1-2 that received EQUIP-T in-service training away from school the previous two years as reported by head teacher/all teachers of Stds 1-2 currently working at the school regardless of standards taught, expressed as a percentage.	Standards 1-2 teachers	
Received school-based EQUIP-T in-service training (% Stds 1-2 teachers):	Number of teachers of Stds 1-2 that received school-based EQUIP-T in-service training the previous two years as reported by head teacher/all teachers of Stds 1-2 currently working at the school regardless of standards taught, expressed as a percentage.	Standards 1-2 teachers	

View of EQUIP-T training (% Stds 1-2 teachers)
    Useful
    Somewhat useful
    Not useful
	The number of teachers of Stds 1-2 that attended that reported they found the EQUIP-T training useful/somewhat useful/not useful/all interviewed teachers of Stds 1-2 who attended EQUIP-T training, expressed as a percentage.	Standards 1-2 teachers	

Gains from EQUIP-T training (% Stds 1-2 teachers):
    Curriculum knowledge
    Subject knowledge
    Teaching skills
    Inclusive teaching skills
    Classroom management/disciplinary skills
    Lesson planning skills
	The number of teachers of Stds 1-2 reporting gain x from the EQUIP-T training/all interviewed teachers of Stds 1-2 who attended EQUIP-T training and thought it was (somewhat) useful, expressed as a percentage.	Standards 1-2 teachers	

Confidence in my teaching			
Support network			
Other			
Proportion of EQUIP-T training attended (% Stds 1-2 teachers):	The number of teachers of Stds 1-2 that report they attended all/most/some of the EQUIP-T school-based training sessions/all interviewed teachers of Stds 1-2 who attended EQUIP-T school-based training, expressed as a percentage.	Standards 1-2 teachers	
What difficulties did you experience with EQUIP-T training (% Stds 1-2 teachers):	The number of teachers of Stds 1-2 reporting difficulty x with the school-based EQUIP-T training/all interviewed teachers of Stds 1-2 who attended EQUIP-T training and thought it was (somewhat) useful, expressed as a percentage.	Standards 1-2 teachers	
Materials difficult			
Too much content			
Limited training time			
Payment insufficient			
Provision of teaching and learning materials (EQUIP-T input)			
Received big and read aloud books 2014-2015 (% schools)	The number of schools that received teacher read aloud books and big books in 2014 and 2015/all schools, expressed as a percentage.	All schools	
Received supplementary readers 2014-2015 (% schools)	The number of schools that received supplementary readers 2014-2015/all schools, expressed as a percentage.	All schools	
Changes in teacher capacity			
Confidence in teaching new Stds 1-2 curriculum (% Stds 1-2 teachers):	The number of teachers of Stds 1-2 that report they feel very confident/fairly confident/not confident teaching the new Stds 1-2 curriculum/all interviewed teachers of Stds 1-2, expressed as a percentage.	Standards 1-2 teachers	
Very confident			
Fairly confident			
Not confident			
Teacher Kiswahili score (% correct answers)	Number of Kiswahili questions answered correctly/total number of Kiswahili questions, expressed as a percentage for each teacher of Standards 1-3 who teach Kiswahili.	Standards 1-3 teachers who teach Kiswahili	
Teacher mathematics score (% correct answers)	Number of mathematics questions answered correctly/ total number of mathematics questions, expressed as a percentage, for each teacher of Standards 1-7 who teach mathematics.	Standards 1-7 teachers who teach mathematics	
Changes in the use of inclusive teaching practices in the classroom (EQUIP-T outcome)			
Teacher interactions with pupils are:			Collection of information: Enumerators observed the entire duration of each and recorded which pupils teachers interacted with noting if the pupil was a boy or girl, and how many boys and girls respectively were present. A classroom gender map was completed for each subject.
Gender balanced			
More with boys			
More with girls	The number of lessons where teachers' interaction with pupils is gender balanced/more with boys/more with girls/all Standards 2 lessons observed, expressed as a percentage. The number of lessons where teachers interacted more with boys/all lessons observed, expressed as a percentage (%). The number of lessons where teachers interacted more with girls/all lessons observed, expressed as a percentage (%).	Observed standard 2 lessons in Kiswahili and mathematics	Indicator construction: First, teacher interactions with girls as a proportion of total teacher interactions with all pupils is computed. Second, the proportion of girls present in the classroom is computed. Teacher interaction is considered gender balanced if the difference between the proportion of interactions with girls and the proportion of girls present during the lesson is smaller than 10 percentage points.

Teacher engaged with at least one pupil from all six areas of the classroom (% all observations)	The number of Std 2 lessons where teacher engaged with at least one pupil from all six areas in the classroom/all Std 2 lessons observed, expressed as a percentage.	Standard 2 lessons observed	Collection of information: A classroom mapping instrument that divides the classroom into six approximately equally-sized areas was used by enumerators to record the number of interactions between teachers and pupils across the six classroom areas.
Distribution of teacher-pupil interactions (% all interactions): Front two areas Middle two areas Back two areas	The number of teacher interactions with pupils in the front two/middle two/bac two area of the classroom/all observed interactions, expressed as a percentage.	Standard 2 lessons observed	Collection of information: a classroom mapping instrument that divides the classroom into six approximately equally-sized areas was used by the enumerator to record the number of interactions between teachers and pupils across the six classroom areas.
Pupils with useable desk space (%)	The number of pupils with useable desk space/the total number of pupils present during the observed Std 2 lesson, expressed as a percentage.	Standard 2 lessons observed	
Pupil absence from school (%)	The proportion of standard 1-3 pupils who were absent from school over all standard 1-3 pupils in school roster, expressed as a percentage	Enrolled standards 1-3 pupils in all schools	Enumerators record all Stds 1-3 classes and count the number of pupils present on the day of the survey. The difference between this head count and the number of Stds 1-3 pupils in the school records are considered the number of absent pupils on the day of survey.
Changes in the use of effective teaching practices in the classroom (EQUIP-T outcome)			
Teaching practices introductory stages of lesson (% lessons): States objectives States new skills to be acquired Checks prior knowledge	The number of observed Std 2 lessons where teachers display teaching practice x fully, partly or not at all during the lesson introductory stages/all Std 2 lesson observations, expressed as a percentage.	Standard 2 lessons observed	For each teaching practice enumerators recorded responses as follows: 'no' if they did not observe the practice, 'partly' if they observed some of parts of the practice and 'yes' if they observed all required aspects of the practice.
Teaching practices concluding stages of lesson (% lessons): Checks acquired new skills Plenary summarising learning	The number of observed Std 2 lessons where teachers display teaching practice x fully, partly or not at all during the lesson concluding stages/all Std 2 lesson observations, expressed as a percentage.		
Teaching practices middle stages of lesson (% lessons): Pupils demonstrate in front of class Teachers asks open ended questions Teacher probes pupil answers Teacher encourages pupil questions Teacher gives feedback on pupil work Teacher uses paired or group work Relates well with and praises pupils Uses different instructional materials	The number of observed Std 2 lessons where teachers display teaching practice x frequently, infrequently or not at all during the middle stages of the lesson/all Std 2 lesson observations, expressed as a percentage.	Standard 2 lessons observed	For each teaching practice enumerators recorded responses as follows: 'no' if practice not observed, 'yes, infrequently' if practice partly observed and 'yes, frequently' if the practice was frequently observed.
Teacher frequently demonstrates: At least seven effective teaching practices (% lessons)	The number of observed Std 2 lessons where teachers frequently demonstrate at least seven out of 14 selected teaching practices/all Std 2 lessons observed, expressed as a percentage.	Standard 2 lessons observed	
Teacher frequently demonstrates: At least three effective teaching practices (% lessons)	The number of observed Std 2 lessons where teachers frequently demonstrate at least three out of 14 selected teaching practices/all Std 2 lessons observed, expressed as a percentage.	Standard 2 lessons observed	

Changes in availability (EQUIP-T output) and use (EQUIP-T outcome) of teaching and learning materials in classrooms			
Received Std 2 supplementary readers in 2014-2015 (% schools)	The number of schools that received supplementary readers for Std 2 in 2014-2015/all schools, expressed as a percentage.	All schools	
Received Std 2 big books in 2014 and 2015 (% schools)	The number of schools that received big books for Std 2 in 2014-2015/all schools, expressed as a percentage.	All schools	
Kiswahili supplementary readers available in classroom (% Kiswahili lessons)	The number of observed Std 2 lessons where Kiswahili supplementary readers are available in the classroom/all observed Kiswahili Std 2 lessons, expressed as a percentage.	Standard 2 Kiswahili lessons observed	
Teacher uses big books (% Std 2 lessons)	The number of Std 2 Kiswahili lessons where teacher used big books/total number of observed Std 2 lessons, expressed as a percentage.	Standard 2 Kiswahili lessons observed	
Pupils read supplementary readers during lessons (% Std 2 lessons)	The number of Std 2 lessons where pupils read supplementary readers to themselves or loud/total number of observed Std 2 lessons, expressed as a percentage.	Standard 2 Kiswahili lessons observed	
Notices groups of pupils with learning difficulties (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 that report noticing groups of pupils in their classes that have learning difficulties/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	
Group of pupils identified to have learning difficulties (% Stds 1-3 teachers):			
With disability or special educational needs			
Boys	The number of teachers of Stds 1-3 that report group x having learning difficulties in their classes/interviewed teachers of Stds 1-3 who reported that they notice groups of pupils with learning difficulties, expressed as a percentage.	Teachers of standards 1-3	
Girls			
Don't speak Kiswahili at home			
Poor pupils			
Haven't attended preschool			
With health problems			
Parents not interested in education			
Other			
No particular group			
Able to help groups of pupils with learning difficulties (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 that report they are able to help pupils with learning difficulties/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	
Action to help pupils with learning difficulties (% Stds 1-3 teachers):			
Adapt materials and teaching to app level			
Use regular assessment to monitor progress			
Ensure pupil engagement in lessons	The number of teachers of Stds 1-3 that report action x to help pupils with learning difficulties/interviewed teachers of Stds 1-3 who reported they are able to help pupils with learning difficulties, expressed as a percentage.	Teachers of standards 1-3	
Give extra tuition classes			
Suggest extra tuition classes by others			
Switch btw Kiswahili and vernacular language			
Talk to pupil's parents			
Group pupils together			

Give more exercises and work			
Repeat topics until pupils understand			
Other			
Teacher reports speaking a language other than Kiswahili at home (%)	The number of teachers of Stds 1-3 reporting they speak a language other than Kiswahili at home/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	
Teacher can speak same local language as pupil (% Std 3 pupils)	Number of Std 3 pupils reporting their teacher can speak the same local language as they/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Teacher reports speaking Kiswahili when teaching (%)	The number of teachers of Stds1-3 reporting they speak Kiswahili when teaching/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	
Teacher speaks Kiswahili during lessons (% Std 3 pupils)	Number of Std 3 pupils reporting the main language used by their teachers during lessons is Kiswahili/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Teacher reports speaking Kiswahili with pupils outside the classroom (%)	The number of Stds 1-3 teachers reporting they speak Kiswahili with pupils outside the classroom/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	
Teacher reports switching between Kiswahili and a vernacular language when teaching (%)	The number of teachers of Stds1-3 reporting they switch between Kiswahili and a vernacular language when teaching/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	
Teacher switches between Kiswahili and a vernacular language during lessons (% Std 3 pupils)	Number of Std 3 pupils reporting their teacher switches between Kiswahili and a vernacular language during lessons/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Teacher switches between Kiswahili and a vernacular language when teaching (% Std 2 lessons)	The number of observed Std 2 lessons where teacher switched between Kiswahili and a vernacular language/all observe Std 2 lessons, expressed as a percentage.	Standard 2 lessons observed	
Teacher reports assessing pupil academic progress during the last five days (%)	The number of teachers of Std 1-3 who report assessing pupil academic progress during the last five days/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standards 1-3 teacher	
Teacher can show example of homework assigned during the last five days (%)	The number of teachers of Stds 1-3 able to show example of homework assigned during the last five days/interviewed teachers of Stds 1-3 that report using pupil assessment, expressed as a percentage.	Standards 1-3 teacher	
Pupil enrollment by Standard (school mean)	The average number of pupils enrolled in Standard x in 2014 and 2016	Head teachers	
Class size by Standard (mean)	The average number of pupils per class in Standard x in 2014 and 2016	Head teachers	
Pupils per classroom in use (school mean)	The average number of pupils (all Standards) per usable classroom in 2014 and 2016	Head teachers	
School has second shift (%)	The number of school that have a second shift/all schools, expressed as a percentage.	Head teachers	
Teachers of Stds 1-7 at same school at BL and ML (%)	The number of teachers of Stds 1-7 that are at the same school at BL and ML/all teachers of Stds 1-7, expressed as a percentage.	Standards 1-7 teachers	From teacher roster.
Reason for leaving for teachers who are no longer at the school at ML (%):	The number of former teachers reported by the head teacher to have left school for reason x/all former teachers, expressed as a percentage.	Former teachers who were at the school at BL but are no longer at the school at ML.	This is head teachers reporting on former teachers.
Transferred			
Disciplinary issue			
Quit job			
Retired			
Passed away			
Long term sick			

Maternity leave			
Went for studies			
Other			
Location of previous posting for teachers who have been at their current school for less than two years (% Stds 1-2 teachers):	The number of teachers of Stds 1-2 reporting previous teaching job in location x/all interviewed teachers of Stds 1-2 who had been teaching at their current school for two years or less, expressed as a percentage.	Standards 1-2 teachers	
Another school in same district			
Another school in same region			
A school in another region			
Near retirement age of 60 years (% Stds 1-2 teachers)	The number of teachers of Stds 1-2 who are 56 years or more at the time of the BL survey and 58 years or more at the time of the ML survey/all interviewed teachers of Stds 1-2, expressed as a percentage.	Standards 1-2 teachers	
Teaching experience (mean years)	The average number of years teachers of Sdts 1-3 have worked as a teacher.	Standards 1-3 teachers	
Age (mean years)	Average teacher age in years.	Standards 1-3 teachers	
Helped at home w homework (% Std 3 pupils)	Number of Std 3 pupils reporting that someone at home helps them with their homework when they need it/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Pupil reads to someone at home (%)	Number of Std 3 pupils reporting that they read to someone at home everyday/sometimes/neve/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Someone at home reads to pupil (%)	Number of Std 3 pupils reporting that someone at home reads to them everyday/sometimes/neve/all assessed Std 3 pupils, expressed as a percentage.	Standard 3 pupils	
Changes in instructional time (EQUIP-T outcome)			
Actual weekly timetabled minutes for mathematics in Stds 1 and 2 (before adjustment).	Minutes per week timetabled for mathematics in Stds 1 and 2 (school mean).	BL: School records – timetables recorded for classes in Stds 1-2. ML: Head teacher interview – timetables recorded for classes in Stds 1-2.	Data on timetables for each class in Stds 1 and 2 were used to identify how many periods by subject are timetabled each week. For each class in a standard, the total number of weekly periods assigned for mathematics and Kiswahili were multiplied by the number of minutes assigned to each period to calculate total weekly minutes in each subject at the class level. These totals were then averaged across the number of classes to get the number of minutes timetabled for each subject by standard. Finally, the weekly minutes were averaged across standards one and two.

To estimate to what extent available instructional time is reduced by classroom absenteeism, indicators on weekly minutes timetabled were adjusted for whether teachers were present in a classroom.

This is a rough estimate of actual instructional time. |
Actual weekly timetabled minutes for mathematics in Stds 1 and 2 after adjusting for the % teachers timetabled to teach and were present in the classroom.	The minutes per week timetabled for mathematics in standards 1 and 2 after adjusting for the percentage of timetabled Sds 1 and 2 teachers in the classroom in the period before lunch (school mean).		
Actual weekly timetabled minutes for Kiswahili in Stds 1 and 2 (before adjustment).	Minutes per week timetabled for Kiswahili in Stds 1 and 2 (school mean).		
Actual weekly timetabled minutes for Kiswahili in Stds 1 and 2 after adjusting for the % of teachers timetabled to teach and present in the classroom.	The minutes per week timetabled for Kiswahili in Stds 1 and 2 after adjusting for the percentage of timetabled Stds 1 and 2 teachers in the classroom in the period before lunch (school mean).		

On the day of the survey of all teachers in the roster: Absent from school (%)	The number of teachers who were not present for the teacher head count on the day of the survey/ all teachers working at the school, expressed as a percentage.	All teachers in schools' teacher rosters	Collection of information: The school and classroom absenteeism measures rely on two different headcounts of teachers carried out by enumerators. At the start of the first day of the school visit, enumerators first recorded teachers who were present at school and second, during the lesson before lunch, recorded if teachers timetabled to teach before lunch were in classrooms teaching.
Of teachers present on the day of the survey and timetabled to teach: Absent from class (%)	The number of teachers who were not present at their timetabled lesson before lunch/all teachers present on the day of the survey and timetabled to teach the lesson before lunch, expressed as a percentage.	Teachers in schools' teacher rosters who were scheduled to teach before lunch and present at school on the day of the survey	Classroom absenteeism was measured during the lesson before lunch because it is a 'typical' lesson time to make the observation that was the same across all surveyed schools, but that avoided the start of the day so that classroom absenteeism was not confounded with lateness.
Of teachers present on the day of the survey: Arrived late (%)	The number of teachers who arrived after the school is officially supposed to start/all teachers present on the day of the survey, expressed as a percentage.	All teachers present in school on day of survey	
Reasons for school absenteeism for teachers who reported being absent from school the last 30 days (% Stds 1-3 teachers): Illness Family responsibility Attending training Official education work Transport problem Collecting salary Other official work Other private work Lack of motivation	The number of teachers of Stds 1-3 who reported being absent from school and reported reason x/all interviewed teachers of Stds 1-3 who reported being absent from school during the last 30 days, expressed as a percentage.	Standards 1-3 teachers	
Reasons for classroom absenteeism for teachers who reported being absent from class the last 30 days (% Stds 1-3 teachers): Large workload Meeting with head teacher Meeting with teachers Lack of motivation Illness Feeling tired	The number of teachers of Stds 1-3 who reported being absent from class and reported reason x/all interviewed teachers of Stds 1-3 who reported being absent from class during the last 30 days, expressed as a percentage.	Standards 1-3 teachers	
Teacher job satisfaction (mean rating)	Mean of self-reported ratings of Stds 1-3 teachers' job satisfaction on the day of the survey.	Standards 1-3 teachers	The rating scale is from one to ten, where 1 indicates 'completely unsatisfied' and ten indicates 'completely satisfied.
Community appreciation (mean rating)	Mean of Stds 1-3 teachers' ratings of how valued they feel by the community on the day of the survey.	Standards 1-3 teachers	The rating scale is from one to ten, where 1 indicates 'completely unsatisfied' and ten indicates 'completely satisfied.
Head teacher appreciation (mean rating)	Mean of Stds 1-3 teachers' ratings of how much they feel their head teacher value them as a teacher on the day of the survey	Standards 1-3 teachers	The rating scale is from one to ten, where 1 indicates 'completely unsatisfied' and ten indicates 'completely satisfied.
Reported teacher satisfaction compared to two years ago (% Stds 1-3 teachers):		Standards 1-3 teachers	

More satisfied	The number of teachers of Stds 1-3 who reported feeling more satisfied/less satisfied/similarly satisfied with their job today than two years ago/all interviewed teachers of Stds 1-3, expressed as a percentage.		
Less satisfied			
Similarly satisfied			
Teacher background characteristics			
Female (% teachers)	Number of teachers that are female/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standards 1-3 teachers	
Highest professional qualification (% Stds 1-3 teachers):	The number of teachers whose highest professional qualification is x/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standards 1-3 teachers	
Bachelors of Education or higher			
Diploma or advanced diploma			
Certificate in education			
Other professional qualification			
No professional qualification			

E.3   Chapter 4: School leadership and management

Indicator name	Indicator definition	Respondent / unit of analysis	Notes
Chapter 4 EQUIP-T component 2: School leadership and management			
Provision of head teacher INSET (EQUIP-T input)			
Attended SLM in-service training last two years (% head teachers)	Number of head teachers that reported attending SLM in-service training the previous two years/all interviewed head teachers, expressed as a percentage.	Head teachers	The relevant period for BL is 2012-2013 and for ML 2014-15.

Attended in-service SLM training provided by (% head teachers):
    EQUIP-T
    LANES
    BRN
    STEP
 	
    Other
	Number of head teachers that reported attending in-service SLM training from provider x the previous two years/interviewed head teachers, expressed as a percentage.	Head teachers	

Attended EQUIP-T SLM in-service training last two years (% head teachers)	Number of head teachers that reported attending EQUIP-T SLM in-service training the previous two years/all interviewed head teachers, expressed as a percentage.	Head teachers	The relevant period for BL is 2012-2013 and for ML 2014-15.
HT´s view of EQUIP-T SLM training (%) Useful Somewhat useful Not useful	The number of head teachers reporting that they found the EQUIP-T SLM training useful/somewhat useful/not useful/interviewed head teachers who attended EQUIP-T SLM training, expressed as a percentage.	Head teachers	
Gains from EQUIP-T SLM training (% head teachers): Knowledge of head teacher responsibilities	The number of head teachers reporting gain x from the EQUIP-T SLM training/interviewed head teachers who attended EQUIP-T SLM training and thought it was (somewhat) useful, expressed as a percentage.	Head teachers	

Teacher management skills			
Financial management skills			
School development planning skills			
Reporting/record keeping skills			
Academic programme management skills			
Confidence in role as head teacher			
Support network			
Stronger relationship with teachers			
Stronger relationships with parents/community			
Other			
What difficulties did head teachers experience with EQUIP-T SLM training (%):			
None			
Not relevant to my job	The number of head teachers reporting difficulty x with the EQUIP-T SLM training/interviewed head teachers who attended EQUIP-T SLM training and thought it was (somewhat) useful, expressed as a percentage.	Head teachers	
Materials difficult			
Too much content			
Too theoretical			
Took too much time			
Payment insufficient			
Other			
Attended Early Grade teaching in-service training last two years (% head teachers)	Number of head teachers that reported attending Early Grade teaching in-service training the previous two years/all interviewed head teachers, expressed as a percentage.	Head teachers	The relevant period for BL is 2012-2013 and for ML 2014-15.
Attended in-service SLM training provided by (% head teachers):			
EQUIP-T			
LANES	Number of head teachers that reported attending in-service Early Grade teaching training from provider x the previous two years/interviewed head teachers, expressed as a percentage.	Head teachers	
BRN			
STEP			
Other			
Changes in head teacher capacity (EQUIP-T output)			
Has WSDP for current school year (% schools)	The number of head teachers reporting they have a whole school development plan (WSDP) for the current school year/all interviewed head teachers, expressed as a percentage.	Head teachers	Head teachers were questioned about whether they had a WSDP for year x. To check the reliability of this response, head teachers were asked to present this WSDP to the interviewer. In addition, interviewers had to review the available WSDPs to identify a range of topics contained within the plans. Topics identified included teaching and learning objectives, a budget and baseline data and objectives.
WSDP comprehensiveness (% schools):			
No WSDP	The number of schools with a WSDP that contains no/one/two/three of the core elements/all schools, expressed as a percentage.	Head teachers	The core elements are: (1) a budget, (2) teaching and learning objectives and (3) baseline data and targets.
Has WSDP but it is not available			
WSDP has none of the core elements			

WSDP has one of the core elements			
WSDP has two of the core elements			
WSDP has three of the core elements			
WSDP contents (% schools):			
Improvements to school facilities			
Teaching and learning objectives	The number of schools with whole school development plans (WSDPs) that contain element x/all schools, expressed as a percentage.	Head teachers	
Strategy to improve Stds 4 and 7 exam scores			
Strategy to reduce dropout or pupil absenteeism			
Strategy to improve girls' learning			
Strategy to improve transition to secondary school			
Budget			
Baseline data and targets			
Changes in head teachers' school leadership and management (EQUIP-T outcome)			
Reported most common teacher performance management practices (% head teachers):			In the head teacher interview, only head teachers were asked this question, not assistant head teachers or academic masters answering on behalf of the head teacher if absent. But some non-interviewed head teachers were phoned for this information to reduce the number of missing observations.
Pupil academic results			
Lesson preparations	The number of head teachers reporting teacher performance management practice x as the most common/all interviewed head teachers, expressed as a percentage.	Head teachers	
Teaching performance in class			
Teacher punctuality and attendance			
Use of continuous pupil assessment			
Other			
Report lesson observation by head teacher (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 who report that the head teacher observes their teaching/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standard 1-3 teachers	
Report written lesson observation feedback from head teacher (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 who report receiving written lesson observation feedback from the head teacher/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standard 1-3 teachers	
Report lesson plans were checked by head teacher (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 who report that the head teacher checks their lesson plans/all interviewed teachers of Stds 1-3 expressed as a percentage.	Standard 1-3 teachers	
Report written lesson plan feedback from head teacher (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 who report receiving written lesson plan feedback from the head teacher/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standard 1-3 teachers	
Report receiving at least one performance appraisal in the previous school year (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 who report that the head teacher, assistant head teacher or academic master held at least one individual meeting with them to discuss their performance and professional development needs during the previous school year/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standard 1-3 teachers	

Four or more staff meetings in the last 60 days (% head teachers)	The number of head teachers responding that at least four staff meetings were held in the last 60 days/all interviewed head teachers, expressed as a percentage.	Head teachers	In the head teacher interview, the CAPI instrument was designed only to ask teacher management related questions of *actual* head teachers (not academic masters or other persons answering on behalf of the head teacher). However, subsequent to the initial survey, head teachers were phoned for this information to reduce the number of missing responses.
Four or more staff meetings in the last 60 days (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 responding that at least four staff meetings were held in the last 60 days/all interviewed teachers of Stds 1-3, expressed as a percentage.	Standard1-3 teachers	
Rewards for teachers who perform well exist (% head teachers)	The number of head teachers reporting that there are rewards in their school for teachers who perform well/all interviewed head teachers, expressed as a percentage.	Head teachers	In the head teacher interview, the CAPI instrument was designed only to ask teacher management related questions of *actual* head teachers (not academic masters or other persons answering on behalf of the head teacher). However, subsequent to the initial survey, head teachers were phoned for this information to reduce the number of missing responses.

Types of teacher performance rewards (% head teachers)
    Financial
    Material (in-kind resources)
    Verbal recognition
    Other
	The number of head teachers reporting reward type x/all interviewed head teachers, expressed as a percentage.	Head teachers	

Rewards for teachers who perform well exist (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 reporting there are rewards in their school for teachers who do well/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	

Types of teacher performance rewards (% Stds 1-3 teachers)
    Financial
    Material (in-kind resources)
    Verbal recognition
    Other
	The number of teachers of Stds 1-3 reporting reward type x/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	

Action is taken for teachers performing poorly (% head teachers)	The number of head teachers reporting that action is taken at their school for teachers who perform poorly/all interviewed head teachers, expressed as a percentage.	Head teachers	

Types of actions for poor teacher performance (% head teachers)
    Extra support to improve teaching
    Increased lesson observation
    Increased checking of lessons plans etc.
    Warning from HT
    HT reports to WEC
    Warning from WEC
    Other
	The number of head teachers reporting action type x/all interviewed head teachers, expressed as a percentage.	Head teachers	

Action is taken for teachers performing poorly (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 reporting that action is taken for teachers who perform poorly/all interviewed teachers of Stds 1-3, expressed as a percentage.	Teachers of standards 1-3	

Types of actions for poor teacher performance (% Stds 1-3 teachers)
    Extra support to improve teaching
    Increased lesson observation
	The number of teachers of Stds 1-3 reporting action type x/all interviewed teachers of Stds 1-3 who reported that head teacher took action to improve education, expressed as a percentage.	Teachers of standards 1-3	

Increased checking of lessons plans etc.			
Warning from HT			
HT reports to WEC			
Warning from WEC			
Other			
Report that head teacher took action to improve education in 2015 (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 who report that the head teacher took action to improve education in 2015/all interviewed teachers of Stds 1-3, expressed as a percentage	Teachers of standards 1-3	
Reported action by head teacher to improve education in 2015 (% Stds 1-3 teachers)	The number of teachers of Stds 1-3 reporting that the head teacher took action x to improve education in 2015/interviewed teachers of Stds 1-3 who reported that the HT took action to improve education, expressed as a percentage.	Teachers of standards 1-3	
Ensuring teacher attendance and punctuality			
Ensuring supply of teaching and learning materials			
Strengthening relationship w parents or community			
Introducing extra tuition classes			
Reducing pupil absenteeism			
Other			
Head teacher at the same school at BL and ML (%)	The number of head teachers that have been head teachers at the same school at BL and ML/all interviewed head teachers, expressed as a percentage.	Head teachers	
Near retirement age of 60 years (% head teachers)	The number of head teachers who are 56 years or more at the time of the BL survey and 58 years or more at the time of the ML survey/all interviewed head teachers, expressed as a percentage.	Head teachers	
Head teachers absent on day of survey using head count observation (%)	The number of head teachers who were not present at the head count on the day of the survey/all head teachers, expressed as a percentage.	Head count	A head count of all head teachers was conducted by enumerators on the day of the survey.
Reasons for school absenteeism for head teachers who report being absent from school the last 30 days (% head teachers):	The number of head teachers who report being absent from school and report reason x for their absenteeism/all interviewed head teachers, expressed as a percentage.	Head teachers	
Illness			
Family responsibility			
Attending training			
Official education work			
Transport problem			
Collecting salary			
Other official work			
Other private work			
Lack of motivation			
Head teacher job satisfaction (mean rating)	Mean of self-reported ratings of head teachers' job satisfaction on the day of the survey.	Head teachers	The rating scale is from one to ten, where 1 indicates 'completely unsatisfied' and ten indicates 'completely satisfied.
Capitation grants			
Estimate of capitation grant payments per pupil received in 2014 (Tsh)	The total amount of capitation grants received in 2014/total number of pupils enrolled in 2014.	Head teachers	

Estimate of capitation grant payments per pupil received in 2015 (Tsh)	The total amount of capitation grants received in 2015/total number of pupils enrolled in 2015.	Head teachers	
Head teacher background			
Female (% head teachers)	Number of female head teachers/all head teachers, expressed as a percentage.	Head teachers	
Age (mean years)	Average head teacher age in years.	Head teachers	
Time working as a head teacher (mean years)	The average number of years head teachers have worked as a head teacher.	Head teachers	
Time working as a HT at current school (mean years)	The average number of years HTs have worked as a HT at their current school.	Head teachers	

Highest professional qualification (% head teachers):
    Bachelors of Education or higher
    Diploma or advanced diploma
    Certificate in education
    Other professional qualification
    No professional qualification
	The number of head teachers whose highest professional qualification is x/all head teachers, expressed as a percentage.	Head teachers	

E.4   Chapter 5 District planning and management

Indicator name	Indicator definition	Respondent/unit of analysis	Notes
Changes in the frequency of school visits by Ward Education Coordinators (WECs) and Quality Assurers (EQUIP-T output)			
Number of WEC visits previous school year (mean)	The average of the total number of visits by WEC to a school during the previous school year as reported by the head teacher (school mean).	Head teachers	
WEC observed teaching during last visit (% schools)	The number of head teachers who reported the WEC observed teaching during last visit/interviewed head teachers who reported receiving at least one WEC visit the previous school year, expressed as a percentage.	Head teachers	
Visited by Quality Assurer previous school year (% schools)	The number of head teachers who report being visited by Quality Assurer in the previous school year/all interviewed head teachers, expressed as a percentage.	Head teachers	QAs were formerly known as School Inspectors.
Changes in WECs capacity (EQUIP-T output) and management practices (EQUIP-T outcome)			
Attended meeting with WEC and other head teachers last 60 days (% head teachers)	The number of head teachers who report they attended a meeting with WEC and other head teachers from the ward the last 60 days/all interviewed head teachers, expressed as a percentage.	Head teachers	

Duration of last WEC visit (% schools):
    30 minutes or less
    1 hour or less
    2 hours or less
    3 hours or less
    More than 3 hours
	The number of head teachers who reported length x/interviewed head teachers who reported receiving at least one WEC visit the previous school year, expressed as a percentage.	Head teachers	

WEC support to school is very good or good (% HTs)	The number of head teachers who stated that the support of the WEC to the school is good or very good/all interviewed head teachers, expressed as a percentage.	Head teachers	
Changes in districts' capacity for planning and management (EQUIP-T output)			

Provides written monthly school reports to WEC/district and able to show them (% schools)	The number of head teachers who provide monthly written school reports to the WEC/district and able to show them/all interviewed head teachers, expressed as a percentage.	Head teachers	

E.5   Chapter 6 Community participation and demand for accountability

Indicator name	Indicator definition	Respondent/unit of analysis	Notes
Provision of training for school committees (SCs) (EQUIP-T input)			
School committee received training (% schools)	The number of head teachers reporting that the school committee received training on its roles and responsibilities in 2014 or 2015/all interviewed head teachers, expressed as a percentage. All schools	Head teachers	
School committee exists (% schools)	The number of head teacher reporting that a school committee exists/all interviewed head teachers, expressed as a percentage.	Head teachers	
School committees' capacity increased (EQUIP-T output)			
Head teachers rating SC support to school as 'very good' or 'good' (% HTs)	The number of head teachers reporting the support of the SC to the school is 'very good' or 'good'/all interviewed head, expressed as a percentage. All head teachers	Head teachers	
Minutes from last School Committee meeting exist (% schools)	The number of schools where the head teacher could show minutes from the last meeting of the SC/all interviewed head teachers, expressed as a percentage.	Head teachers	
Main topics discussed at last SC meeting (% schools): 　Academic progress 　Pupil discipline 　Pupil absenteeism/dropout 　Teacher discipline 　Teacher supervision/support 　School development plan 　School finance including parental contributions 　Infrastructure development 　Other	The number of schools where the last meeting of the SC covered topic x/all interviewed head teachers, expressed as a percentage.	Head teachers	
Formation of parent teacher partnerships (PTPs) (EQUIP-T input)			
School has a PTP (% schools)	The number of head teachers reporting that a PTP exists/all interviewed head teachers, expressed as a percentage.	Head teachers	
PTP received training (% schools)	The number of head teachers reporting that the PTG received training on its roles and responsibilities in 2014 or 2015/all interviewed head teachers, expressed as a percentage.	Head teachers	
Provider of PTP training (% schools): 　EQUIP-T 　3Rs (LANES) 　BRN 　TZ21 　STEP 　UNICEF Other	The number of PTPs receiving training on its roles and responsibilities from provider x in 2014 or 2015/all interviewed head teachers, expressed as a percentage.	Head teachers	

PTP took action to improve education in the school in 2015 (% schools)	The number of head teachers reporting that the PTP took action to improve education in the school in 2015/all interviewed head teachers, expressed as a percentage.	Head teacher	
Action taken by PTP to improve education in the school in 2015 reported by HT ensure pupil attendance (% schools):	The number of schools where PTP took action to ensure pupil attendance in 2015/all interviewed head teachers, expressed as a percentage.	Head teacher	
Community needs assessments developed (EQUIP-T input) and feed into school development plans (EQUIP-T output)			
Community carried out community needs assessment (% schools)	The number of head teachers reporting that the community carried out its own needs assessment and wrote it down in 2014 or 2015/all interviewed head teachers, expressed as a percentage.	Head teachers	
Action was taken by school or community based on community needs assessment (% schools)	The number of head teachers reporting that the school or the community took action based on the community needs assessment in 2014 or 2015/all interviewed head teachers, expressed as a percentage.	Head teachers	.
School/community action taken based on community needs assessment: improve school infrastructure (% schools)	The number of head teachers reporting that the school or the community improved school infrastructure based on the community needs assessment in 2014 or 2015/all interviewed head teachers, expressed as a percentage.	Head teachers	
Notice boards display relevant information publicly (EQUIP-T input) and improved communication between schools and communities (EQUIP-T outcome)			
Schools with notice board publicly displayed on school premises (% schools)	The number of schools that have a notice board displayed publicly displayed on school premises/total number of schools, expressed as a percentage.	Schools	
Types of info displayed on school notice board (% schools): Plan/financial info Academic info Attendance info Events info	The number of schools that display information x/all schools, expressed as a percentage.	Schools	
Head teacher holds at least one meeting per year with teachers and all parents (% schools)	The number of head teacher reporting that they held at least one meeting with teachers and all parents last year/all interviewed head teachers, expressed as a percentage.	Head teachers	This means a meeting where all parents are invited, not a meeting of the parent teacher partnership (PTP).
Head teachers rating community support to the school as 'very good' or 'good' (% HTs)	The number of head teachers reporting the support of the community to the school is 'very good' or 'good'/all interviewed head teachers interviewed, expressed as a percentage.	Head teachers	
Main topics discussed at last teacher and all parents meeting (% schools): Academic progress Pupil discipline Pupil absenteeism/dropout	The number of schools where the last meeting of the teacher and *all* parents meeting covered topic x/total number of interviewed head teachers that that reported holding a teachers and *all* parents meeting last year, expressed as a percentage.	Head teachers	This means a meeting where all parents are invited, not a meeting of the parent teacher partnership (PTP).

Teacher discipline			
Teacher supervision/support			
School development plan			
School finance incl. parental contributions			
Infrastructure development			
Other			

Annex F   Detailed statistical tables of results from programme treatment districts

F.1   Pupil learning and background characteristics

Annex table 5 Pupil learning and background characteristics in programme treatment schools

Indicator[1]	BL Estimate	BL P10	BL P90	BL SE	BL Lower 95CI	BL Upper 95CI	BL N	ML Estimate	ML P10	ML P90	ML SE	ML Lower 95CI	ML Upper 95CI	ML N
Kiswahili														
Pupil ability score (logits, mean)	-1.6	-4.5	0.4	0.1	-1.8	-1.3	1,487	-0.7***	-3.5	0.6	0.1	-0.9	-0.5	1,463
Distribution of pupils by band (% of pupils)														
Pupils in band 0(%)	39.4			2.7	34.0	44.8	1,487	23.2***			2.5	18.2	28.1	1,463
Pupils in band 1E(%)	8.0			0.9	6.2	9.9	1,487	6.4			0.9	4.6	8.2	1,463
Pupils in band 1A(%)	16.8			1.3	14.3	19.3	1,487	19.6			1.5	16.5	22.6	1,463
Pupils in band 2E(%)	23.8			1.9	19.9	27.6	1,487	28.5*			1.7	25.0	31.9	1,463
Pupils in band 2A(%)	12.1			1.6	9.0	15.2	1,487	22.4***			2.2	18.1	26.8	1,463
Maths														
Pupil ability score (logits, mean)	-1.0	-3.4	1.4	0.1	-1.3	-0.8	1,495	-0.6**	-3.3	1.6	0.1	-0.8	-0.4	1,483
Distribution of pupils by band (% of pupils)														
Pupils in band 0(%)	13.2			1.4	10.4	16.1	1,495	11.3			1.4	8.5	14.1	1,483
Pupils in band 1E(%)	27.8			1.9	24.1	31.6	1,495	19.5***			1.6	16.4	22.7	1,483
Pupils in band 1A(%)	30.7			1.7	27.2	34.1	1,495	31.9			1.5	28.9	34.8	1,483
Pupils in band 2E(%)	23.9			2.0	19.9	27.9	1,495	30.4**			1.8	26.7	34.0	1,483
Pupils in band 2A(%)	4.4			0.8	2.9	6.0	1,495	7.0*			1.1	4.9	9.1	1,483
Background characteristics: support at home														
Helped with homework as needed (% Std 3 pupils)	.	.	.	.	.	.	.	62.7			2.5	57.7	67.8	1,329
Pupil reads to someone at home (% Std 3 pupils)								.	.	.	.	.	.	.
Never								21.5			2.2	17.2	25.9	1,264
Sometimes								69.0			2.3	64.3	73.6	1,264
Everyday								9.5			1.2	7.2	11.8	1,264
Someone reads to pupil at home (% Std 3 pupils)								.	.	.	.	.	.	.
Never								34.0			2.3	29.5	38.5	1,263
Sometimes								59.5			2.3	54.8	64.1	1,263
Everyday								6.5			0.9	4.6	8.4	1,263
Language support at school														

Teachers speaks same language as pupil (% Std 3 pupils)	.	.	.	.	.	.	.	21.3			2.1	17.0	25.5	1,116
Teacher switches language during lessons (% Std 3 pupils)	.	.	.	.	.	.	.	16.1			1.2	13.7	18.5	1,473

Sources: IE ML survey. Note: (1) Blank cells mean that the estimate is not applicable to this type of indicator. (2) Asterisks indicate statistical significance levels *** p<0.01, ** p<0.05, * p<0.1

Annex table 6 Mean pupil scores in Kiswahili and maths at ML by background characteristics

	Boys	Girls	Poorer	Richer	Kiswahili/Foreign	Local Language
Kiswahili						
Pupil ability score (logits, mean)	-0.8**	-0.6	-0.9**	-0.6	-0.2***	-0.8
N	723.0	740.0	531.0	925.0	317.0	1,141.0
Maths						
Pupil ability score (logits, mean)	-0.5**	-0.7	-0.9**	-0.5	0.2***	-0.9
N	734.0	749.0	536.0	940.0	320.0	1,158.0

Sources: IE ML survey. Note: (1) Asterisks indicate statistical significance levels *** p<0.01, ** p<0.05, * p<0.1

F.2   Teachers capacity, performance and conditions for learning

Annex table 7 Teachers' capacity performance and conditions for learning in programme treatment schools

Indicator[1]	BL Estimate	BL P10	BL P90	BL SE	BL Lower 95CI	BL Upper 95CI	BL N	ML Estimate	ML P10	ML P90	ML SE	ML Lower 95CI	ML Upper 95CI	ML N
Personal characteristics														
Female (% Stds 1-3 teachers)	55.6			3.9	48.0	63.2	327	58.4			3.1	52.3	64.5	384
Age (mean years)	39.6	24.0	57.0	1.0	37.6	41.6	327	37.5	25.0	58.0	0.9	35.8	39.2	384
Work experience and tenure														
Time working as a teacher (mean years)	15.8	1.0	35.0	1.0	13.8	17.8	327	13.8	2.0	35.0	0.9	12.0	15.6	384
Time teaching at current school (mean years)	8.4	1.0	23.0	0.6	7.2	9.5	327	7.8	1.0	22.0	0.5	6.7	8.8	384
Near retirement age 60 years (% teachers)	18.6			2.9	12.8	24.4	327	10.1**			1.8	6.6	13.6	384
Highest professional qualification (% Stds 1-3 teachers)														
Bachelors of Education or higher (% teachers)	0.4			0.5	-0.6	1.4	326	0.2			0.2	-0.3	0.6	384
Diploma or advanced diploma (% teachers)	1.5			0.9	-0.3	3.3	326	1.8			0.8	0.2	3.4	384
Certificate in education (% teachers)	94.2			2.1	90.1	98.3	326	96.0			1.2	93.6	98.3	384
Other professional qualification (% teachers)	3.3			2.0	-0.6	7.3	326	1.3			0.7	-0.1	2.7	384

No professional qualification (% teachers)	0.5			0.6	-0.7	1.8	326	0.7			0.4	-0.1	1.5	384
Satisfaction ratings: Teachers of Stds 1-3 self-reported														
Job satisfaction (mean rating)	7.7	5	10	0.2	7.4	8.1	327	7.9	5	10	0.2	7.6	8.2	341
Community appreciation (mean rating)	6.5	2	10	0.3	6	7	324	6.1	3	10	0.2	5.6	6.6	339
Head teacher appreciation (mean rating)	8.6	6	10	0.1	8.3	8.8	326	8.4	6	10	0.1	8.2	8.6	335
More satisfied	.	.	.	.	.	.	.	55.4			3.7	48.0	62.7	349
Less satisfied	.	.	.	.	.	.	.	29.2			3.8	21.7	36.7	349
Similarly satisfied	.	.	.	.	.	.	.	15.5			2.1	11.3	19.6	349
Teachers' use of assessment														
Teacher does not use regular pupil assessment (% Stds 1-3 teachers)	30.3			4.0	22.3	38.2	327	42.5**			3.6	35.4	49.6	341
Teacher gives homework (% Stds 1-3 teachers who use pupil assessment)	6.4			3.2	0.2	12.7	229	9.9			3.4	3.1	16.6	201
Language use														
Speaks Swahili when teaching (% Stds 1-3 teachers)	.	.	.	.	.	.	.	100.0			0.0	100.0	100.0	382
Switch language when teaching (% Stds 1-3 teachers)	.	.	.	.	.	.	.	26.0			3.5	19.0	33.0	382
Speaks Swahili outside of the classroom (% Stds 1-3 teachers)	.	.	.	.	.	.	.	92.4			1.9	88.6	96.3	382
Speaks Kiswahili at home (% Stds 1-3 teachers)	.	.	.	.	.	.	.	90.3			1.9	86.6	94.1	384
Teacher switches language (% of lessons)	4.0			1.4	1.2	6.7	199	3.7			1.4	0.9	6.5	231
Reasons for school absenteeism (% Stds 1-3 teachers that reported being absent from school last 30 days)														
Illness	.	.	.	.	.	.	.	40.8			4.5	31.7	50.0	289
Attending training	.	.	.	.	.	.	.	19.1			3.3	12.4	25.7	289
Official educ work	.	.	.	.	.	.	.	16.9			2.9	11.1	22.6	289
Collecting salary	.	.	.	.	.	.	.	15.5			2.9	9.7	21.3	289
Family responsibility	.	.	.	.	.	.	.	15.4			3.3	8.8	22.1	289
Other private work	.	.	.	.	.	.	.	9.3			2.6	4.2	14.5	289
Other official work	.	.	.	.	.	.	.	5.5			1.4	2.6	8.3	289
Lack of motivation, salary	.	.	.	.	.	.	.	0.8			0.7	-0.6	2.2	289
Lack of motivation, housing	.	.	.	.	.	.	.	0.3			0.4	-0.4	1.1	289
Lack of motivation, other	.	.	.	.	.	.	.	0.0			0.0	0.0	0.0	289
Transport problem	.	.	.	.	.	.	.	0.3			0.3	-0.2	0.8	289
Other (%)	.	.	.	.	.	.	.	10.1			3.2	3.7	16.6	289
Reasons for class absenteeism (% Stds 1-3 teachers that reported being absent from the classroom last 30 days)														
Large workload	.	.	.	.	.	.	.	29.9			6.3	17.0	42.8	106
Illness	.	.	.	.	.	.	.	23.1			5.4	12.2	34.1	106

Meeting with other teachers	.	.	.	.	.	.	.	16.5			5.3	5.8	27.3	106
Meeting with head teacher	.	.	.	.	.	.	.	12.0			4.0	3.8	20.2	106
Lack of motivation, housing, salary and other	.	.	.	.	.	.	.	0.0			0.0	0.0	0.0	106
Other	.	.	.	.	.	.	.	33.7			6.0	21.4	46.0	106
Supporting conditions: salary and time to school														
Time to school (mean minutes)								14.5	0.0	40.0	1.5	11.5	17.6	384
Last three salares received in full (% Stds 1-3 teachers)								93.3			2.5	88.4	98.3	376
Last three salaries received on time(% Stds 1-3 teachers)								97.1			1.4	94.3	99.8	379
Teaching practices: Teacher interaction with pupils is (% all interactions)														
Gender balanced	53.8			4.3	45.2	62.3	193	64.7*			3.7	57.4	72.1	225
More with boys	30.3			3.8	22.9	37.7	193	23.9			3.1	17.8	30.0	225
More with girls	16.0			3.2	9.6	22.3	193	11.4			2.6	6.2	16.5	225
At least one pupil from all six areas of the classroom	58.5			4.0	50.7	66.3	199	79.4***			2.5	74.4	84.3	231
Distribution of teacher-pupil interactions (% of interactions)														
Front two areas	41.5	23.5	66.7	1.5	38.4	44.5	193	38.3*	24.6	50.0	0.9	36.4	40.2	225
Middle two areas	30.4	8.3	47.6	1.5	27.4	33.3	193	34.1**	24.1	42.9	0.8	32.6	35.7	225
Back two areas	28.2	10.3	45.5	1.2	25.8	30.6	193	27.6	13.6	38.9	0.8	25.9	29.2	225
Teaching practices: Teaching behaviours during the introductory lesson stages (% lessons)														
States objectives of lesson	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes	23.0			3.1	16.9	29.1	199	9.9***			2.0	6.0	13.8	231
Partly	48.7			4.0	40.8	56.6	199	34.4***			3.7	27.2	41.6	231
No	28.3			3.6	21.2	35.3	199	55.7***			3.6	48.6	62.9	231
States new skills to be acquired	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes	7.2			1.8	3.6	10.8	199	10.1			2.2	5.7	14.4	231
Partly	40.5			3.9	32.7	48.3	199	25.4***			3.4	18.8	32.0	231
No	52.3			3.9	44.7	59.9	199	64.5**			3.5	57.7	71.3	231
Checks prior knowledge	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes	21.5			2.8	16.1	27.0	199	45.8***			3.6	38.7	53.0	231
Partly	38.4			3.5	31.6	45.2	199	18.5***			2.7	13.2	23.8	231
No	40.1			3.4	33.3	46.8	199	35.6			3.2	29.3	42.0	231
Teaching practices: Teaching behaviours during the concluding lesson stages (% lessons)														
Checks pupils have acquired new skills	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes	21.8			2.9	16.1	27.5	199	23.1			3.1	17.0	29.2	231
Partly	47.9			4.0	40.0	55.9	199	23.3***			2.6	18.2	28.5	231
No	30.3			3.5	23.4	37.1	199	53.6***			2.9	47.8	59.4	231
Holds plenary summarising learning	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes	21.1			2.6	16.1	26.1	199	13.1**			2.3	8.6	17.5	231

Partly	41.5			3.4	34.8	48.3	199	20.7***			2.9	14.9	26.4	231
No	37.4			3.1	31.3	43.5	199	66.3***			2.7	60.9	71.6	231
Teaching practices: Teaching behaviours during the middle lesson stages (% lessons)														
Pupils demonstrate for class	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	35.8			3.6	28.8	42.9	199	43.1			3.6	35.9	50.2	231
Yes, infrequently	27.5			3.4	20.9	34.1	199	22.8			3.5	15.9	29.6	231
No	36.7			3.5	29.7	43.7	199	34.2			3.1	28.1	40.3	231
Asks open-ended questions	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	11.2			2.3	6.7	15.7	199	11.3			2.4	6.6	16.1	231
Yes, infrequently	23.2			3.6	16.0	30.3	199	8.2***			2.2	3.9	12.4	231
No	65.6			3.4	58.9	72.4	199	80.5***			2.7	75.2	85.8	231
Probes pupil answers (%)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	11.8			1.8	8.2	15.4	199	7.0*			1.7	3.8	10.3	231
Yes, infrequently	37.0			3.4	30.4	43.7	199	34.9			3.3	28.4	41.4	231
No	51.2			3.3	44.6	57.8	199	58.0			3.1	51.9	64.2	231
Encourages pupil questions	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	4.0			1.8	0.4	7.6	199	1.1			0.7	-0.2	2.4	231
Yes, infrequently	16.7			2.5	11.8	21.5	199	24.2*			3.4	17.6	30.9	231
No	79.3			2.9	73.5	85.1	199	74.7			3.4	68.0	81.3	231
Gives feedback on pupil work	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	25.7			3.4	19.0	32.5	199	46.9***			3.7	39.5	54.2	231
Yes, infrequently	32.5			3.7	25.1	39.9	199	19.1***			2.8	13.5	24.6	231
No	41.8			3.5	35.0	48.6	199	34.1			3.4	27.4	40.7	231
Uses paired or group work	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	6.5			1.7	3.2	9.8	199	5.7			1.8	2.2	9.2	231
Yes, infrequently	16.6			3.3	10.2	23.0	199	18.6			3.1	12.5	24.8	231
No	77.0			3.2	70.6	83.3	199	75.6			3.5	68.9	82.4	231
Praises pupils				.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	51.0			3.6	43.9	58.2	199	48.1			3.2	41.8	54.4	231
Yes, infrequently	33.8			3.6	26.8	40.8	199	47.3***			3.3	40.8	53.7	231
No	15.2			2.9	9.5	20.9	199	4.6***			1.1	2.4	6.8	231
Uses instructional materials (%)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yes, frequently	18.3			3.0	12.4	24.2	199	13.0			2.4	8.3	17.7	231
Yes, infrequently	26.0			3.5	19.1	32.9	199	40.0***			3.6	32.9	47.1	231
No	55.6			3.6	48.6	62.7	199	47.0*			3.7	39.7	54.4	231
Teaching practices: Teacher frequently demonstrated (% of lessons)														
At least seven effective teaching practices	9.0			1.8	5.4	12.5	199	9.5			2.1	5.4	13.7	231
At least three effective teaching practices	57.7			3.5	50.9	64.5	199	60.1			3.4	53.5	66.8	231
Teaching practices: Use of resources														
Teacher used big books (% of lessons)	.	.	.	.	.	.	.	7.0			2.0	2.9	11.1	137

Teaching practices: Number of swahili supplementary readers available in the classroom												
None	.	.	.	.	.	.	.	88.2			2.8	82.7	93.7	137
1 to 20	.	.	.	.	.	.	.	8.5			2.2	4.1	12.9	137
21 to 50	.	.	.	.	.	.	.	0.5			0.6	-0.7	1.7	137
More than 50	.	.	.	.	.	.	.	2.8			1.5	-0.2	5.7	137
Teaching practices: Pupils read supplementary readers (% of Kiswahili lessons)												
Most pupils	.	.	.	.	.	.	.	3.7			1.8	0.0	7.3	137
Some pupils	.	.	.	.	.	.	.	3.6			1.9	-0.1	7.3	137
No pupils	.	.	.	.	.	.	.	92.7			2.3	88.2	97.2	137
Teaching: Pupils with useable desk space (% lessons)												
Pupils with useable desk space	71.7		2.3	67.2	76.3	199	77.1*			1.9	73.4	80.8	231
Teaching practices: Inclusion												
Notices pupils with learning difficulties (% Stds 1-3 teachers)							99.5			0.3	98.8	100.1	384
Teaching practices: Group identified as having learning difficulties (% Stds 1-3 teachers):												
Do not speak Kiswahili at home							43.2			3.3	36.5	49.8	380
Poor							32.0			3.0	25.9	38.1	380
Parents uninterested							25.3			2.8	19.6	30.9	380
Girls							20.8			3.0	14.9	26.8	380
Boys							17.9			3.2	11.6	24.3	380
Have not attended preschool							10.8			1.9	7.0	14.6	380
Disability							8.2			2.6	2.9	13.4	380
Health problems							5.3			1.5	2.4	8.2	380
Able to help pupils with learning difficulties (% Stds 1-3 teachers)							98.5			0.9	96.8	100.3	379
Teaching practices: Action taken to help pupils with learning difficulties (% Stds 1-3 teachers):												
Give extra tuition							56.7			3.9	48.8	64.6	375
Group pupils together							29.1			3.2	22.7	35.6	375
Talk to parents							24.1			2.6	18.8	29.4	375
Ensuring pupil engagement							18.3			2.8	12.6	23.9	375
Switch language							12.5			2.3	8.0	17.0	375
Repeat topics							10.0			1.8	6.4	13.5	375
Regular assessment							9.7			1.9	5.8	13.6	375
Adapt materials/teaching							3.7			1.1	1.5	5.8	375
Other							13.2			3.1	7.0	19.3	375
Teaching capacity: Teacher confidence in teaching new curriculum (% of Stds 1-2 teachers)												
Very confident							74.4			4.3	65.8	83.0	220
Fairly confident							24.2			4.3	15.7	32.8	220
Not confident							1.3			1.3	-1.3	4.0	220
Teacher management												
Four or more staff meetings last 60 days (% Stds 1-3 teachers)	15.4		2.9	9.8	21.1	327	31.6***			3.7	24.2	39.0	384

Report lesson plan checked by HT (% teachers of Stds 1-3)	91.1			1.9	87.3	95.0	327	93.1			1.6	89.9	96.3	341
Report written lesson plan feedback from HT (% teachers of Stds 1-3)	47.3			4.3	38.8	55.7	327	23.4***			3.6	16.3	30.5	341
Report lesson observation by HT (% teachers of Stds 1-3)	52.4			4.0	44.4	60.4	325	38.8**			3.7	31.5	46.1	341
Report written lesson observation feedback (% teachers of Stds 1-3)	4.6			1.5	1.6	7.7	325	4.7			1.7	1.4	8.0	340
Report performance appraisal (% teachers of Stds 1-3)	27.7			3.8	20.0	35.0	327	29.4			3.8	21.7	37.0	341
Teacher performance rewards exist (% teachers of Stds 1-3)								13.9			2.6	8.7	19.2	376
Teacher management: Teacher performance reward types (% teachers of Stds 1-3)														
Financial								29.3			9.7	8.6	50.0	65
Material								28.0			7.9	10.9	45.0	65
Verbal recognition								42.4			10.5	19.8	64.9	65
Certificate/cup/medal								15.0			5.3	3.5	26.4	65
Action taken for poor teacher performance (% teachers of Stds 1-3)								66.6			3.5	59.5	73.6	361
Teacher management: Types of action taken for poor teacher peformance (% teachers of Stds 1-3):														
Extra teaching support provided								8.6			2.2	4.1	13.1	241
More head teacher lesson observations								7.5			1.9	3.8	11.3	241
Head teachers check lesson plans								5.8			2.3	1.2	10.5	241
Warning from head teacher								88.8			3.4	82.0	95.6	241
Head teacher reports to WEC								18.8			4.1	10.7	27.0	241
Warning from WEC								15.5			4.3	6.9	24.1	241
Warning from academic master/teacher								8.2			2.8	2.6	13.8	241
Head teacher took action to improve education quality in 2015 (% Stds 1-3 teachers)								74.9			3.6	67.7	82.1	362
Teacher management: Reported head teacher action to improve education quality in 2015 (% Stds 1-3 teachers):														
Ensuring teacher attendance and punctuality								36.8			3.4	29.9	43.7	273
Introducing extra tuition classes								35.0			3.8	27.3	42.7	273
Ensuring supply of teaching and learning materials								25.9			3.1	19.7	32.1	273
Actions to reduce pupil absenteeism								18.2			3.6	11.0	25.4	273
Strengthening relationship with parents/community								16.5			2.7	11.2	21.8	273
Teacher INSET: Of interviewed Stds 1-2 teachers (%)														

Attended any training	12.0		3.5	5.1	18.9	180	95.5***			1.4	92.8	98.2	218
Attended EQUIP-T	0.0		0.0	0.0	0.0	180	95.3***			1.4	92.6	98.0	218
Attended LANES	0.0		0.0	0.0	0.0	180	0.0			0.0	0.0	0.0	218
Attended BRN	4.8		2.7	-0.6	10.2	180	0.2*			0.0	0.2	0.3	218
Attended STEP	0.0		0.0	0.0	0.0	180	0.0			0.0	0.0	0.0	218
Attended OTHER	8.1		3.3	1.6	14.6	180	8.7			3.0	2.7	14.7	218
Teacher INSET: Teachers' view of EQUIP-T training (% Stds 1-2 teachers who attended)													
Useful							95.4			1.7	92.0	98.8	201
Somewhat useful							4.6			1.7	1.2	8.0	201
Not useful							0.0			0.0	0.0	0.0	201
Teacher INSET: Gains from EQUIP-T INSET (% Stds 1-2 teachers who attended)													
Curriculum knowledge							53.5			5.5	42.4	64.5	201
Subject knowledge							32.5			4.5	23.4	41.6	201
Teaching skills							78.1			4.2	69.8	86.5	201
Inclusive teaching skills							54.6			4.4	45.6	63.5	201
Classroom management skills							20.0			3.0	14.1	26.0	201
Lesson planning skills							27.7			4.4	18.8	36.5	201
Confidence in own teaching							27.5			4.3	18.9	36.0	201
Support network							3.6			2.5	-1.5	8.6	201
Other							16.0			3.5	8.9	23.1	201
Teacher INSET: Difficulties with EQUIP-T training (% Stds 1-2 teachers who attended)													
None							43.5			4.0	35.6	51.4	201
Materials difficult							9.2			2.2	4.7	13.6	201
Too much content							12.3			2.5	7.3	17.2	201
Payment insufficient							16.7			2.9	10.9	22.6	201
Limited training time							15.0			3.8	7.4	22.5	201
Teacher INSET: Attendance, Reported amount of EQUIP-T school-based training attended (% Stds 1-2 teachers who attended)													
All							65.9			4.7	56.4	75.3	181
Most							15.6			3.8	7.9	23.3	181
Some							18.5			3.7	11.1	25.9	181
Attended away and school-based EQUIP-T training (% Stds 1-2 teachers who attended)							89.1			2.4	84.4	93.8	198
Teacher turnover													
Teacher no longer at the school at ML (% BL Stds 1-7 teachers, unweighted)	.	.	.	.	.	.	.	31.5			1.4	28.7	34.3	1,047
Teacher turnover: Leaving reasons for teachers who left school since BL (% of Stds 1-7 teacher who left, unweighted estimates):													
Transferred	.	.	.	.	.	.	.	56.7			2.7	51.3	62.0	330
Disciplinary issue	.	.	.	.	.	.	.	0.9			0.5	-0.1	1.9	330
Quit job	.	.	.	.	.	.	.	1.5			0.7	0.2	2.8	330
Retired	.	.	.	.	.	.	.	12.7			1.8	9.1	16.3	330
Passed away	.	.	.	.	.	.	.	2.4			0.8	0.8	4.1	330

Indicator	BL Estimate	BL P10	BL P90	BL SE	BL Lower 95CI	BL Upper 95CI	BL N	ML Estimate	ML P10	ML P90	ML SE	ML Lower 95CI	ML Upper 95CI	ML N
Long term sick	.	.	.	.	.	.	.	0.9			0.5	-0.1	1.9	330
Maternity leave	.	.	.	.	.	.	.	1.5			0.7	0.2	2.8	330
Went for studies	.	.	.	.	.	.	.	22.1			2.3	17.6	26.6	330
Other	.	.	.	.	.	.	.	1.2			0.6	0.0	2.4	330
Teacher turnover: Location of previous teaching job (% of Stds 1-2 teachers who joined school since BL)														
Another school in this district	.	.	.	.	.	.	.	40.6			11.2	17.3	63.8	23
Another school in this region	.	.	.	.	.	.	.	46.5			12.5	20.5	72.5	23
Another school in another region	.	.	.	.	.	.	.	12.9			7.2	-2.1	27.9	23
Teacher subject knowledge: Kiswahili subject knowledge of Stds 1-3 Kiswahili teachers														
Questions correct (%)	58.2	39.5	74.4	1.1	56.0	60.4	247	60.4	41.9	76.7	1.3	57.7	63.0	239
Stds 1-4 qns correct (%)	66.4	42.9	85.7	1.5	63.4	69.4	247	68.9	47.6	85.7	1.5	66.0	71.9	240
Stds 5-7 qns correct (%)	50.4	36.4	68.2	1.0	48.3	52.5	247	52.2	36.4	72.7	1.5	49.2	55.2	239
Reading comprehension qns correct (%)	68.3	50.0	80.0	1.1	66.1	70.6	247	69.2	50.0	80.0	1.4	66.5	71.9	240
Grammatical and punctuation qns correct (%)	42.1	14.3	71.4	1.7	38.7	45.4	247	43.7	14.3	78.6	2.1	39.6	47.9	239
Direct and indirect tenses qns correct (%)	68.6	44.4	88.9	1.6	65.5	71.8	247	72.2	55.6	88.9	1.5	69.2	75.1	240
Synonyms and proverbs qns correct (%)	61.3	30.0	90.0	1.6	58.2	64.5	247	64.2	40.0	90.0	1.5	61.3	67.2	240
Teacher subject knowledge: Mathematics subject knowledge of Stds 1-7 mathematics teachers														
Questions correct (%)	59.0	27.8	86.1	1.3	56.5	61.6	506	61.9	30.6	88.9	1.5	58.9	64.9	470
Stds 1-3 qns correct (%)	88.1	66.7	100.0	1.1	86.0	90.3	506	86.7	66.7	100.0	1.5	83.7	89.8	470
Stds 4-5 qns correct (%)	62.3	25.0	87.5	1.3	59.8	64.9	506	62.9	25.0	87.5	1.6	59.7	66.2	470
Stds 6-7 qns correct (%)	54.5	20.0	88.0	1.6	51.4	57.6	506	58.6*	24.0	88.0	1.6	55.4	61.7	470
Whole number qns correct (%)	69.0	42.9	85.7	1.2	66.6	71.3	506	69.3	42.9	100.0	1.6	66.2	72.4	470
Fraction, decimal and percent qns correct (%)	66.7	27.3	90.9	1.1	64.6	68.9	506	71.8***	27.3	100.0	1.5	68.8	74.7	470
Geometry qns correct (%)	43.1	0.0	83.3	1.9	39.2	47.0	506	42.3	0.0	83.3	2.4	37.5	47.1	470
Statistics qns correct (%)	65.9	20.0	100.0	1.9	62.1	69.6	506	71.2*	40.0	100.0	1.9	67.4	75.1	470
Algebra qns correct (%)	45.4	0.0	100.0	2.9	39.6	51.2	506	48.9	0.0	100.0	2.3	44.4	53.4	470

Sources: IE ML survey. Note: (1) Asterisks indicate statistical significance levels *** p<0.01, ** p<0.05, * p<0.1

F.3   School leadership and management

Annex table 8 School leadership and management in programme treatment schools

Indicator[1]	BL Estimate	BL P10	BL P90	BL SE	BL Lower 95CI	BL Upper 95CI	BL N	ML Estimate	ML P10	ML P90	ML SE	ML Lower 95CI	ML Upper 95CI	ML N
Personal characteristics														

Female (% HTs)	15.8			4.8	6.2	25.4	99	17.3			4.9	7.6	27.1	99
Age (mean years)	43.5	31.0	56.0	1.1	41.2	45.7	98	41.9	29.0	58.0	1.2	39.4	44.4	98
Time working as HT (mean years)	.	.	.	.	.	.	.	7.1	0.0	18.0	1.0	5.1	9.2	96
Time working as head teacher at current school (mean years)	4.0	0.0	9.0	0.6	2.8	5.2	99	4.1	0.0	10.0	0.5	3.0	5.2	99
Highest professional qualification (% of HTs):														
Bachelors of Education or higher	0.0			0.0	0.0	0.0	99	4.4*			2.3	-0.3	9.0	99
Diploma or advanced diploma	8.4			2.6	3.1	13.6	99	16.8			5.2	6.4	27.3	99
Certificate in education	90.5			2.9	84.7	96.4	99	78.8**			5.6	67.5	90.1	99
Other professional qualification	1.1			1.3	-1.6	3.7	99	0.0			0.0	0.0	0.0	99
No professional qualification	0.0			0.0	0.0	0.0	99	0.0			0.0	0.0	0.0	99
Absenteeism														
Head teachers absent on the day of the survey using head count observations (%)	16.2			4.0	8.2	24.2	100	14.6			4.9	4.8	24.5	100
Report absence from school in last 30 days (% HTs)								99.6			0.5	98.6	100.6	93
Absenteeism: Reasons for absence from school (%HT)														
Did not report absence								0.4			0.5	-0.6	1.4	93
Illness								7.0			2.8	1.4	12.6	93
Family responsibilities								11.2			3.2	4.7	17.8	93
Attending training								32.6			5.8	21.0	44.2	93
Official education work								77.3			5.0	67.3	87.2	93
Transport								0.0			0.0	0.0	0.0	93
Collecting salary								11.7			4.2	3.2	20.2	93
Other official work								29.7			5.3	19.1	40.3	93
Other private work								4.1			3.3	-2.6	10.7	93
Motivation related to housing, salary and other								0.0			0.0	0.0	0.0	93
Other								10.6			5.8	-1.2	22.3	93
HT turnover														
Head teacher at same school at BL and ML (% HTs)	.	.	.	.	.	.	.	53.5			6.2	41.0	66.0	100
HT near retirement age 60	15.3			4.2	6.8	23.8	98	10.0			3.7	2.7	17.4	98
HT INSET														
HT attended any SLM training (%, overlapping BL and ML sample)	12.0			5.1	1.7	22.3	86	70.9***			5.8	59.3	82.6	86
HT attended SLM training by (%, full ML sample)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EQUIP	.	.	.	.	.	.	.	66.5			5.5	55.5	77.5	93
LANES	.	.	.	.	.	.	.	0.9			1.2	-1.4	3.3	93
BRN	.	.	.	.	.	.	.	8.3			4.6	-1.0	17.5	93
STEP	.	.	.	.	.	.	.	0.0			0.0	0.0	0.0	93
OTHER	.	.	.	.	.	.	.	1.7			1.1	-0.4	3.9	93

Did not attend any SLM training	.	.	.	.	.	.	.	31.7			5.6	20.3	43.0	93
HT attended any early-grade teaching training (%)	.	.	.	.	.	.	.	89.0			2.8	83.3	94.7	93
HT attended early grade teaching training by (%)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EQUIP	.	.	.	.	.	.	.	81.4			5.1	71.1	91.6	93
LANES	.	.	.	.	.	.	.	2.8			3.4	-4.1	9.7	93
BRN	.	.	.	.	.	.	.	10.6			3.4	3.8	17.5	93
STEP	.	.	.	.	.	.	.	2.2			1.6	-1.1	5.5	93
OTHER	.	.	.	.	.	.	.	3.6			1.6	0.3	6.8	93
HT did not attend early grade teaching training	.	.	.	.	.	.	.	11.0			2.8	5.3	16.7	93
Head teachers' view of EQUIP SLM training (% HTs attending)								.	.	.	.	.	.	.
Useful								97.7			3.2	91.1	104.3	63
Somewhat useful								2.3			3.2	-4.3	8.9	63
Not useful								0.0			0.0	0.0	0.0	63
Skills gained during EQUIP-T SLM training (% HTs attending)								.	.	.	.	.	.	.
Knowledge HT responsibilities								71.7			9.6	52.1	91.3	63
Teacher mgt skills								73.4			7.2	58.6	88.2	63
Financial mgt skills								29.7			9.3	10.7	48.7	63
School development planning skills								51.9			8.6	34.3	69.6	63
Reporting/record keeping skills								30.1			9.7	10.3	50.0	63
Support network								14.7			5.8	2.8	26.6	63
Stronger relationship with teachers								22.9			8.3	5.9	39.8	63
Stronger relation with parents and community								27.6			8.9	9.4	45.7	63
Difficulties with EQUIP-T SLM training (% HTs attending)								.	.	.	.	.	.	.
None								34.9			9.6	15.2	54.6	63
Not relevant to my job								0.0			0.0	0.0	0.0	63
Materials difficult								2.6			1.9	-1.3	6.5	63
Too much content								38.7			10.1	18.0	59.3	63
Payment insufficient								27.3			8.2	10.6	44.0	63
Limited training time								3.4			2.1	-0.9	7.7	63
Transport difficult or venue far away								7.9			3.7	0.4	15.3	63
Other								12.7			7.1	-1.8	27.2	63
WSDPs														
Has WSDP (% schools)	36.4			5.2	26.0	46.7	100	67.7***			4.6	58.5	76.9	100
WSDP available (% schools)	20.6			3.5	13.6	27.6	100	57.4***			5.4	46.7	68.2	100

WSDP comprehensiveness (% schools)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
No WSDP	63.6			5.2	53.3	74.0	100	32.3***			4.6	23.1	41.5	100
Has WSDP but not available	15.8			4.2	7.4	24.1	100	10.3			3.6	3.0	17.6	100
None	6.1			2.3	1.4	10.8	100	15.0*			3.6	7.7	22.3	100
One core element	7.2			2.5	2.2	12.3	100	28.3***			4.7	18.8	37.8	100
Two core elements	5.0			1.9	1.1	8.9	100	10.7			4.2	2.3	19.1	100
Three core elements	2.2			1.5	-0.8	5.3	100	3.4			1.8	-0.3	7.1	100
WSDP contents (% schools)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Improvements to school facilities	15.2			4.4	6.5	23.9	100	49.9***			5.4	39.1	60.7	100
Teaching and learning objectives	10.8			2.8	5.1	16.4	100	30.3***			5.3	19.7	41.0	100
How to improve Stds 4 and 7 exam scores	7.4			2.6	2.1	12.6	100	15.3*			4.2	6.9	23.8	100
Pupil dropout/absenteeism	6.3			2.1	2.0	10.5	100	18.9**			5.3	8.3	29.5	100
Girls' learning	1.0			1.1	-1.3	3.3	100	11.9**			4.2	3.6	20.3	100
Secondary school transition	4.5			1.9	0.8	8.2	100	6.3			2.4	1.4	11.2	100
Budget	9.3			2.4	4.6	14.1	100	22.9**			5.3	12.2	33.5	100
Baseline data and targets	3.9			1.8	0.2	7.6	100	8.8			2.6	3.6	14.1	100
Other	.	.	.	.	.	.	.	12.8			4.3	4.2	21.4	100
Teacher management														
Head teachers most commonly reported teacher performance management practices (% of HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Monitoring of pupil academic results	17.9			5.8	6.2	29.6	85	21.5			6.4	8.4	34.5	85
Observe lesson preparation	34.0			6.3	21.4	46.7	85	27.1			5.8	15.5	38.8	85
Observe teacher performance in class	24.4			7.1	10.1	38.6	85	30.8			6.8	17.1	44.6	85
Monitoring of teacher punctuality/attendance	4.4			2.4	-0.4	9.2	85	5.0			3.3	-1.6	11.6	85
Ues of continuous pupil assessment	8.2			2.8	2.5	13.9	85	10.5			3.8	2.9	18.2	85
Other	11.1			5.4	0.1	22.1	85	5.0			4.3	-3.6	13.6	85
Reasons for teacher absence from school (% of HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Motivation related to housing, salary level and other	.	.	.	.	.	.	.	12.4			3.0	2.6	14.5	93
Reasons teachers absent from class (% of HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Large workload	.	.	.	.	.	.	.	38.5			6.8	24.9	52.2	93
Teacher performance rewards exist (% of HTs)	33.2			7.7	17.7	48.7	86	46.3			7.1	31.9	60.7	86
Types of rewards (% of HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Financial	13.9			5.4	3.0	24.9	86	9.0			3.1	2.7	15.2	86
Verbal recognition	15.8			6.2	3.2	28.4	86	31.5**			6.5	18.4	44.6	86

Trips/events	3.5			3.4	-3.4	10.3	86	0.0			0.0	0.0	0.0	86
Poor teacher performance action exists (% of HTs)	.	.	.	.	.	.	.	81.3			5.7	69.8	92.9	93
Types of actions for poor teacher peformance (% of HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Extra teaching support is provided	.	.	.	.	.	.	.	11.6			5.1	1.3	22.0	93
More lesson observations	.	.	.	.	.	.	.	6.2			4.3	-2.6	14.9	93
Head teachers check lesson plans	.	.	.	.	.	.	.	2.0			1.2	-0.4	4.5	93
Warning from head teacher	.	.	.	.	.	.	.	58.1			6.1	45.9	70.3	93
Head teacher reports to WEC	.	.	.	.	.	.	.	3.5			2.2	-1.0	8.0	93
Warning from WEC	.	.	.	.	.	.	.	6.9			4.3	-1.8	15.6	93
Head teacher meets and advices teacher	.	.	.	.	.	.	.	18.1			5.5	7.0	29.2	93
Four or more staff meetings last 60 days (% HTs)	23.7			6.1	11.4	35.9	86	22.9			6.6	9.5	36.3	86
Has assigned teacher INSET coordinator (% schools)	.	.	.	.	.	.	.	98.3			1.5	95.3	101.4	93
School physical facilities														
Schools with teacher housing (%)	.	.	.	.	.	.	.	87.2			4.2	78.8	95.5	100
Functional toilet (% schools)	95.8			0.8	94.2	97.3	100	97.4			2.2	93.1	101.8	100
Available drinking water (% schools)	31.9			5.2	21.6	42.3	100	34.8			5.8	23.2	46.4	100
Functional electricity (% schools)	4.5			2.4	-0.2	9.3	100	3.7			1.6	0.5	7.0	100
Staff room (% schools)	86.1			4.0	78.1	94.1	100	97.7**			2.1	93.4	101.9	100
School library (% schools)	12.5			2.4	7.7	17.2	100	14.0			3.4	7.1	20.8	100
Working computers (% schools)	0.9			1.0	-1.2	2.9	100	2.3			1.3	-0.4	5.0	100
Schools with second shift (%)	47.9			6.0	35.9	59.8	100	66.5**			5.5	55.4	77.6	100
Pupil absenteeism on the day of the survey from head count (% Stds 1-3 pupils)														
All pupils	33.5	15.0	53.6	1.7	30.2	36.8	97	25.2***	11.3	44.0	1.6	22.0	28.4	97
Boys	34.6	16.0	57.9	1.9	30.8	38.4	97	26.1***	9.3	44.4	1.6	22.9	29.3	97
Girls	32.4	13.4	49.1	1.7	29.0	35.8	97	24.4***	8.6	41.2	1.7	21.0	27.8	97
Pupil enrolment														
Preschool pupils per school (mean)	56.2	0.0	130.0	6.2	43.6	68.7	87	90.8***	7.0	178.0	7.0	76.7	104.9	87
Pre-school pupils per class (mean)	63.6	16.0	129.0	5.9	51.6	75.6	70	101.9***	28.5	178.0	7.7	86.1	117.8	70
From school records	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Std 1 pupils per school (mean)	83.7	45.0	138.0	4.1	75.6	91.9	95	116.1***	45.0	196.0	8.5	99.1	133.1	95
Std 1 pupils per class (mean)	75.1	42.3	122.0	3.6	67.8	82.3	94	97.5***	41.0	152.0	5.8	85.8	109.1	94
Std 2 pupils per school (mean)	72.8	31.0	128.0	4.0	64.8	80.7	100	78.2	29.0	141.0	5.3	67.6	88.9	100
Std 2 pupils per class (mean)	64.6	25.0	106.0	3.4	57.8	71.4	100	71.5	29.0	119.0	5.0	61.5	81.5	100
Std 3 pupils per school (mean)	65.2	28.0	113.0	3.6	58.0	72.4	99	63.9	24.0	111.0	3.8	56.2	71.6	99
Std 3 pupils per class (mean)	59.9	28.0	100.0	3.3	53.3	66.4	99	61.9	24.0	108.0	3.7	54.5	69.3	99

Indicator	BL Estimate	BL P10	BL P90	BL SE	BL Lower 95CI	BL Upper 95CI	BL N	ML Estimate	ML P10	ML P90	ML SE	ML Lower 95CI	ML Upper 95CI	ML N
Stds 1-7 pupils per school (mean)	486.4	233.0	836.0	21.7	442.8	530.0	100	472.3	231.0	810.0	22.0	428.3	516.4	100
Stds 1-7 pupils per teacher (mean)	53.9	30.8	82.6	2.6	48.7	59.1	100	51.0	28.8	84.4	2.8	45.3	56.7	100
Stds 1-7 pupils per class (mean)	62.7	33.3	92.9	2.7	57.4	68.1	99	63.2	33.0	100.1	2.8	57.6	68.8	99
Stds 1-7 pupils per clasrom in use (mean)								73.7*	38.5	118.6	3.6	66.6	80.9	100
Instructional time														
Official guidelines minutes mathematics Stds 1-2	210.0							240.0						
Weekly minutes Stds 1-2 mathematics before adjustment	214.8	165.0	285.0	6.8	201.1	228.5	98	227.9	195.0	240.0	4.2	219.5	236.4	98
Weekly minutes Std 1-2 mathematics after adjustment	91.0	69.9	120.7	2.9	85.2	96.8	98	144.8***	123.8	152.4	2.7	139.4	150.1	98
Official guidelines minutes Kiswahili Stds 1-2	180.0							480.0						
Weekly minutes Stds 1-2 Kiswahili before adjustment	203.5	150.0	330.0	8.6	186.2	220.7	95	455.1***	405.0	480.0	7.8	439.5	470.7	95
Weekly minutes Stds 1-2 Kiswahili after adjustment	86.2	63.5	139.8	3.6	78.9	93.5	95	289.0***	257.2	304.9	4.9	279.1	299.0	95
Resources received (% schools)														
Supplementary reading books	.	.	.	.	.	.	.	77.2			4.9	67.4	87.0	99
Big books	.	.	.	.	.	.	.	88.5			4.6	79.4	97.7	99

Sources: IE ML survey. Note: (1) Asterisks indicate statistical significance levels *** p<0.01, ** p<0.05, * p<0.1

F.4 District planning and management

Annex table 9 District planning and management in support of programme treatment schools

Indicator[1]	BL Estimate	BL P10	BL P90	BL SE	BL Lower 95CI	BL Upper 95CI	BL N	ML Estimate	ML P10	ML P90	ML SE	ML Lower 95CI	ML Upper 95CI	ML N
WEC support to schools														
School visit by WEC (% schools)	99.1			0.0	99.0	99.2	100	98.7			0.7	97.2	100.1	100
Number of WEC visits (mean)	6.6	2.0	12.0	0.7	5.1	8.0	100	11.5***	4.0	20.0	0.9	9.8	13.2	100
School visit by DSI (% schools)	63.4			5.7	52.0	74.7	100	54.9			6.5	41.8	67.9	100
Number of DSI visits (mean)	1.0	0.0	2.0	0.1	0.7	1.2	100	1.0	0.0	3.0	0.1	0.7	1.2	100
Expected no of WEC visits (% HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
More than once a week	.	.	.	.	.	.	.	9.6			4.8	-0.1	19.3	81
Once per week	.	.	.	.	.	.	.	27.7			6.0	15.6	39.8	81
Several times per month	.	.	.	.	.	.	.	43.8			7.9	27.9	59.8	81
Once per month	.	.	.	.	.	.	.	13.2			4.5	4.0	22.4	81
Several times per year	.	.	.	.	.	.	.	2.6			2.2	-1.8	6.9	81
Don't know	.	.	.	.	.	.	.	3.1			3.8	-4.6	10.9	81

Duration last WEC visit (% HTs)	.	.	.	.	.	.	.	.			.	.	.	.
No WEC visit	.	.	.	.	.	.	.	0.7			0.0	0.7	0.8	100
30 mins or less	.	.	.	.	.	.	.	16.5			4.8	7.0	26.1	100
1 hour or less	.	.	.	.	.	.	.	25.7			5.8	14.2	37.3	100
2 hours or less	.	.	.	.	.	.	.	23.9			6.4	11.0	36.7	100
3 hours or less	.	.	.	.	.	.	.	7.6			3.2	1.2	14.1	100
More than 3 hours	.	.	.	.	.	.	.	24.6			5.9	12.9	36.4	100
Don't know	.	.	.	.	.	.	.	0.8			0.7	-0.6	2.2	100
WEC action in last visit (% HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
No WEC visit	.	.	.	.	.	.	.	0.7			0.0	0.7	0.8	100
Admin	.	.	.	.	.	.	.	57.4			6.4	44.6	70.2	100
Advice on school	.	.	.	.	.	.	.	0.6			0.7	-0.8	2.0	100
Observed lessons	.	.	.	.	.	.	.	12.5			4.0	4.4	20.6	100
Observed facilities	.	.	.	.	.	.	.	8.6			3.6	1.3	15.8	100
Observed mgmt prac	.	.	.	.	.	.	.	12.3			5.2	1.9	22.7	100
Observed school comm	.	.	.	.	.	.	.	1.0			0.1	0.9	1.1	100
Observed PTG	.	.	.	.	.	.	.	2.6			1.3	0.1	5.2	100
Attended INSET	.	.	.	.	.	.	.	4.9			3.9	-3.0	12.7	100
Bringing & supervising exams	.	.	.	.	.	.	.	10.3			3.3	3.7	17.0	100
Coaching & participate in sports	.	.	.	.	.	.	.	1.8			1.6	-1.3	4.9	100
Other	.	.	.	.	.	.	.	11.0			3.8	3.3	18.7	100
How helpful was WEC visit (%HTs)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
No WEC visit	.	.	.	.	.	.	.	0.9			0.1	0.8	1.0	82
Very helpful	.	.	.	.	.	.	.	74.7			5.9	62.9	86.6	82
Fairly helpful	.	.	.	.	.	.	.	23.7			5.9	11.9	35.6	82
Not helpful	.	.	.	.	.	.	.	0.6			0.8	-1.0	2.3	82
Refused	.	.	.	.	.	.	.	0.0			0.0	0.0	0.0	82
WEC support is v good/good (%HT)	79.0	0.0	100.0	4.7	69.4	88.6	82	91.3			4.0	83.2	99.3	82
Attended meeting WEC/HT (%HT)	.	.	.	.	.	.	.	91.3			4.3	82.6	100.1	82
Information flow between schools and districts														
Monthly school report available (% schools)								100.0			0.0	100.0	100.0	97
Report available for (% schools)								.	.	.	.	.	.	.
Jan 2016								75.7			5.3	65.0	86.5	97
Feb 2016								80.7			4.7	71.2	90.1	97
Mar 2016								83.0			4.1	74.7	91.3	97

F.5 Community participation and demand for accountability

Annex table 10 Community participation and demand for accountability in programme treatment schools

Indicator[1]	BL Estimate	BL P10	BL P90	BL SE	BL Lower 95CI	BL Upper 95CI	BL N	ML Estimate	ML P10	ML P90	ML SE	ML Lower 95CI	ML Upper 95CI	ML N
School committees														
School committee exists(%SCH)	99.5			0.6	98.2	100.7	100	98.5			1.8	94.8	102.1	100
Minutes from SC exist(%SCH)	91.4			3.8	83.7	99.0	100	84.2			3.9	76.4	92.0	100
Main topic discussed (%SCH)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
No school committee	0.5			0.6	-0.7	1.8	100	1.5			1.8	-2.1	5.2	100
Academic progress	13.1			3.9	5.3	20.9	100	10.9			4.0	2.7	19.0	100
Pupil discipline/absent	18.4			5.6	7.1	29.7	100	7.5			3.4	0.6	14.3	100
Teacher discipline	1.3			1.1	-1.0	3.6	100	0.0			0.0	0.0	0.0	100
Teacher supervision/support	0.0			0.0	0.0	0.0	100	0.6			0.7	-0.8	2.0	100
School development plan	7.4			2.5	2.3	12.5	100	15.4			5.0	5.5	25.3	100
School finance	25.6			6.3	13.0	38.2	100	40.1			5.8	28.6	51.7	100
Infrastructure	15.8			5.5	4.8	26.7	100	14.7			3.2	8.3	21.1	100
Other	17.9			5.0	7.9	27.8	100	8.3			3.8	0.7	15.8	100
Don't know	0.0			0.0	0.0	0.0	100	1.1***			0.1	0.9	1.2	100
Parent-Teacher Groups (called Parent-Teacher Partnerships at ML)														
School has PTG (%SCH)	14.0			4.1	5.7	22.3	100	96.1***			2.5	91.1	101.2	100
PTG took some action (%SCH)	.	.	.	.	.	.	.	46.9			6.5	33.9	59.9	100
Action taken by PTG (%SCH)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
No PTG	.	.	.	.	.	.	.	3.9			2.5	-1.2	8.9	100
PTG did not take any action	.	.	.	.	.	.	.	49.2			6.5	36.1	62.3	100
Monitor tch attend/punct	.	.	.	.	.	.	.	4.4			2.4	-0.4	9.3	100
Ensure pupil attend/punct	.	.	.	.	.	.	.	30.8			5.8	19.2	42.4	100
Assist in classroom	.	.	.	.	.	.	.	0.3			0.4	-0.4	1.1	100
Provide teach/learn mat	.	.	.	.	.	.	.	0.6			0.7	-0.8	1.9	100
Improve Sch infrastructure	.	.	.	.	.	.	.	7.6			3.7	0.1	15.1	100
Organise School feeding	.	.	.	.	.	.	.	1.4			1.4	-1.3	4.1	100
Organise IGAs	.	.	.	.	.	.	.	1.7			2.0	-2.4	5.8	100
Organise Pupil counsel/welfare	.	.	.	.	.	.	.	3.0			3.5	-4.0	9.9	100
Fundraising	.	.	.	.	.	.	.	0.9			0.6	-0.2	2.0	100
Organise extra tuition	.	.	.	.	.	.	.	1.6			1.0	-0.3	3.5	100
Organise extra curricular	.	.	.	.	.	.	.	2.4			1.9	-1.5	6.3	100
Other	.	.	.	.	.	.	.	10.8			4.8	1.2	20.4	100
Community-led school needs assessments														
Comm needs assess (%SCH)	.	.	.	.	.	.	.	43.7			7.2	29.2	58.2	95

Indicator	Baseline	SE	CI low	CI high	N	Midline	SE	CI low	CI high	N
Action taken needs assess (%SCH)	.	.	.	.	.	32.8	7.2	18.3	47.2	95
Type of action taken (%SCH)	.	.	.	.	.	.	.	.	.	.
No community assessment	.	.	.	.	.	56.3	7.2	41.8	70.8	95
No action taken	.	.	.	.	.	10.9	5.4	0.1	21.8	95
Monitor tch attend/punct	.	.	.	.	.	0.0	0.0	0.0	0.0	95
Ensure pupil attend/punct	.	.	.	.	.	5.7	4.3	-2.9	14.3	95
Assist in classroom	.	.	.	.	.	0.0	0.0	0.0	0.0	95
Provide teach/learn mat	.	.	.	.	.	1.0	0.4	0.1	1.8	95
Improve Sch infrastructure	.	.	.	.	.	22.9	5.8	11.1	34.6	95
Organise School feeding	.	.	.	.	.	7.0	4.9	-2.8	16.9	95
Organise IGAs	.	.	.	.	.	0.8	1.1	-1.3	2.9	95
Organise Pupil counsel/welfare	.	.	.	.	.	0.0	0.0	0.0	0.0	95
Fundraising	.	.	.	.	.	3.1	3.4	-3.7	9.8	95
Organise extra tuition	.	.	.	.	.	0.0	0.0	0.0	0.0	95
Organise extra curricular	.	.	.	.	.	0.0	0.0	0.0	0.0	95
Other	.	.	.	.	.	1.5	0.4	0.7	2.3	95
HT rates SC v good/good (%HT)	53.6	6.1	41.2	65.9	84	74.2**	6.4	61.2	87.1	84
HT rates comm v good/good (%HT)	2.8	0.1	2.5	3.0	82	3.2***	0.1	2.9	3.5	82
Parents and teachers (PT) meetings										
PT meet once a year (%SCH)	86.7	3.1	80.5	92.8	100	96.2***	1.6	93.1	99.3	100
Minutes from PT exist (%SCH)	67.4	5.2	57.0	77.9	100	66.3	4.7	56.9	75.8	100
Main topics discussed (%SCH)	.	.	.	.	.	.	.	.	.	.
No PT meeting	13.3	3.1	7.2	19.5	100	3.8***	1.6	0.7	6.9	100
Academic progress	27.8	5.6	16.6	39.0	100	32.4	6.1	20.2	44.6	100
Pupil discipline/absent	22.3	5.5	11.4	33.2	100	19.3	5.5	8.3	30.3	100
Teacher discipline	0.0	0.0	0.0	0.0	100	0.8	0.9	-1.1	2.6	100
Teacher supervision/support	0.6	0.7	-0.8	2.0	100	1.2	1.1	-1.0	3.4	100
School development plan	8.2	4.0	0.2	16.1	100	6.6	2.6	1.4	11.8	100
School finance	0.0	0.0	0.0	0.0	100	11.7**	5.3	1.0	22.4	100
Infrastructure	5.4	2.6	0.1	10.7	100	9.3	3.1	3.0	15.6	100
School feeding program	.	.	.	.	.	4.6	2.8	-1.1	10.3	100
School committee	.	.	.	.	.	0.5	0.6	-0.7	1.6	100
Other	22.4	4.3	13.7	31.1	100	10.2**	3.7	2.9	17.6	100
SC and PTP training										
SC received training (%SCH)						76.2	4.3	67.4	84.9	94
Provider (%SCH)						.	.	.	.	.
No School committee						1.6	2.0	-2.4	5.6	94
SC did not receive training						22.2	4.2	13.8	30.6	94
EQUIP-T						43.5	5.9	31.7	55.4	94
Ward Education Coordinator						24.6	5.7	13.1	36.0	94
Head teacher						2.5	1.8	-1.2	6.2	94
Other						5.6	1.9	1.7	9.5	94

PTG received training (%SCH)							41.5			6.4	28.6	54.4	96
Provider (%SCH)							.	.	.	.	.	.	.
No PTG							4.0			2.7	-1.5	9.6	96
PTG did not receive training							54.5			6.6	41.2	67.7	96
EQUIP-T							24.4			5.2	14.1	34.8	96
Ward Education Coordinator							6.9			4.1	-1.3	15.1	96
Head teacher							8.3			4.7	-1.2	17.9	96
WEC and HT							1.9			1.2	-0.6	4.4	96
Other							0.0			0.0	0.0	0.0	96
School and community communication													
Schools with notice board (%SCH)	49.0			6.8	35.4	62.5	100	71.7***			5.7	60.3	83.0	100
Types of info displayed (%SCH)	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Plan/finanical info	1.5			1.1	-0.7	3.8	100	31.9***			5.6	20.6	43.2	100
Academic info	10.3			4.3	1.7	19.0	100	32.8***			5.8	21.2	44.4	100
Pupil/tch attendance info	8.2			3.3	1.7	14.8	100	6.5			2.6	1.3	11.8	100
Sch/community events info	7.7			3.0	1.7	13.8	100	18.1*			5.1	7.9	28.2	100

Sources: IE ML survey. Note: (1) Asterisks indicate statistical significance levels *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Annex G   Measurement of pupil learning outcomes

G.1   Summary of the content of the pupil tests

G.1.1 Rationale for using EGRA- and EGMA- type tests and matching to curriculum criteria

As explained in Chapter 3 which covers the quantitative IE design, the BL pupil tests were adapted from existing EGRA and EGMA instruments. These are competency-focused instruments. Part of the decision to use these types of instruments for the EQUIP-T measurement of pupil learning was because at the time of the BL IE survey, the Government had recently used EGRA and EGMA instruments in a survey to monitor its then flagship national education programmes BRN-Ed, and as a BL for another national education programme LANES. The Government was in the process of setting national targets related to the results of these tests.

As set out in the IE BL report (OPM 2015b, pp100,103), the skills tested in the IE tests were matched, as far as possible, to the 'specific objectives' laid out in the existing Standards 1 and 2 Kiswahili and mathematics curricula (MoEVT, 2005a,b) which explained what the pupil should be able to do to reach the curriculum standard (e.g. Standard 2 pupils should be able to add numbers to get a sum not exceeding 1000). Two tables (one each for Kiswahili and maths) set out the list of skills that pupils had to demonstrate in the IE BL tests to be considered as achieving in one of five curriculum-linked performance band (OPM 2015b, pp102,104). The competencies required to move up the scale are in a logical order of increasing difficulty and it was noted that 'these [competencies] appear to be broadly consistent with the order of the competencies expressed in the Standard 1 and 2 curriculum, although in many cases the curriculum statements are fairly general and similar at the two levels' (OPM 2015b, p100).

Subsequent to the BL IE research, the Government rolled out a new Standards 1 and 2 curriculum in 2015 which focuses on the 3Rs competencies of reading, writing and arithmetic (MoEVT 2016). Another EGRA and EGMA survey was carried out to continue the monitoring of the Government's BRN-Ed and LANES programmes (RTI 2016).

Because of the content and nature of the BL IE tests, there was no need to adapt them to fit with the new Standards 1 and 2 curriculum. The same pupil tests were used in the BL and ML IE. This also has the advantage of making the raw-score results from traditional test analysis comparable over time (see Section G.2 below). A similar exercise of mapping the skills tested in the IE tests to the 'competence benchmarks' in the new curricula found that the BL classification of skills into performance bands is still valid. Of course, the same limitations as were acknowledged at BL also apply at ML: (i) many of the 3Rs curriculum statements are fairly general (e.g. Standard 1 pupils should be able to read aloud with appropriate speed) and so cannot be mapped with precision; and (ii) the IE tests do not cover all of the competencies listed in the new 3Rs curriculum (the skills that are covered are discussed next). Annex table 12 and Annex table 13 in this annex show the mappings.

G.1.2 Kiswahili

Skill areas: There are seven subtests in total. Each subtest covers a different skill area:

- Four subtests are timed oral reading tests of syllables, familiar words, invented words and reading a short passage;

- The remaining three subtests cover: reading comprehension (five questions), listening comprehension (five questions), and writing/spelling dictated sentences (two sentences).

Curriculum levels: the short passage was designed to be a Standard 2 level text and so the reading comprehension questions which relate to this are Standard 2 level questions.[53] The remaining subtests combine Standard 1 and Standard 2 curriculum skills by including questions of different levels within each subtest. The 3Rs Standard 2 curriculum requires that pupils read text with accuracy and fluency and, although this is not specified in the curriculum itself, the Government has set a national target for reading speed of 50 words per minute for Standard 2 pupils.[54]

G.1.3 Mathematics

Skill areas: There are six subtests containing 60 questions in total. These cover: number comparison/quantity discrimination (eight questions), missing numbers in sequences (eight questions), addition (16 questions), subtraction (16 questions), multiplication (8 questions), and word problems (4 questions).

Curriculum levels: Apart from multiplication, the other five subtests contain a mix of Standard 1 and Standard 2 level questions. Multiplication was part of the previous Standard 2 level curriculum, but it is not part of the new 3Rs curriculum for Standard 2. For comparability with BL, the multiplication subtest was retained in the ML test. Another reason is that Standard 3 pupils would most likely have been taught the old Standard 2 curriculum for the first half of 2015 (since the new curriculum was rolled out during 2015) so may well have learnt multiplication. Indeed Standard 3 pupils' raw multiplication scores increased significantly between BL and ML (see Chapter 7). Over the whole test, the balance is skewed towards Standard 1 level material; about 60% of the questions are at the lower curriculum level.

G.2 Notes on traditional test analysis in the IE

This part of the pupil test analysis relies on simple descriptive statistics of the different subtest results, such as mean test scores, mean reading speeds, and the proportion of pupils achieving more than x% of questions correct. These supplementary results are in Chapter 7 Section 7.1. In interpreting these results, it is important to understand how the subtests were marked, and how non-response was treated.

Marking of the Kiswahili subtests: The four reading subtests are 'marked' using a simple reading speed indicator: number of words correctly read per minute. Each pupil was given exactly one minute to complete each reading test. If a pupil finished early, this was accounted for in the reading speed. For the remaining subtests, marks are awarded as follows: reading comprehension (five marks: one per question); listening comprehension (five marks: one per question); writing (21 marks for spelling words and punctuation).

Marking of the maths subtests: One mark is given for each question answered correctly. The number of questions in each subtest is given above.

Treatment of non-response: 'Non-response' is treated as incorrect on all subtests in the traditional test analysis, except the four reading speed subtests in Kiswahili because this does not

---

[53] The reading passage was developed by a team of experienced Tanzanian subject and test design specialists (see Chapter 0 for more details).
[54] Well-known international research (Abadzi, 2006) found that reading at 45-60 words per minute is a minimum fluency required for comprehension.

affect the 'reading speed' indicator. Most non-response happened because of instructions in the test to skip questions to enhance the efficiency of the subtest when a pupil got a fixed number of prior questions incorrect. The test designers sought to make the questions in each subtest hierarchically difficult. In Kiswahili, for example, the writing subtests contained two sentences, if the pupil was unable to write any word correctly in the first sentence, then the second sentence was skipped. In mathematics, for example, the addition and subtraction questions were divided into two levels, with level two questions designed to be more difficult than level one questions. If a pupil did not get any level one questions correct (one and two digit problems) then level two questions (two and three digit problems) were skipped. Given this hierarchical ordering of questions within the subtests, it was deemed reasonable to treat the skipped questions as incorrect, as it is very unlikely that the students who were unable to answer the less difficult items correctly would have been able to answer the more difficult items correctly if they were administered to them.

G.3   Application of the Rasch model in the IE

This subsection explains the rationale for using Rasch modelling to analyse the Kiswahili and maths test data for the IE. It discusses the principles underpinning the Rasch model and some of its key assumptions.

The key advantage of using Rasch modelling to analyse pupil test scores for the IE, is that, under certain assumptions, this generates estimates of pupil 'ability' in Kiswahili and mathematics on an *interval scale* which can be directly linked to criterion-referenced competencies found in the curriculum. On an interval scale, equal differences between numbers (in this case, pupil ability estimates) reflect equal differences in the amount of the underlying attribute being measured. Since the key objective of the IE is to measure change in learning achievement over time, an interval measurement scale allows for more accurate estimation of change. Using raw scores and traditional test analysis for this purpose can be substantially misleading (Wright and Stone 1979).

A key principle underlying the Rasch model is that of seeking to measure a latent unidimensional trait. This simply means an underlying construct that cannot be measured directly and can be thought of in terms of more or less. The IE seeks to measure the latent unidimensional traits of literacy skills (in Kiswahili) and numeracy (a type of mathematical) skills.

The Rasch model is the simplest Item Response Theory (IRT) model. It is a probabilistic mathematical model of a person's (in our case, a pupil's) response to an item whereby, relative to an item of a certain difficulty, as a pupil's level of ability (as estimated across all items) increases, the probability of a correct response increases. The latent trait is conceived as a single dimension along which items can be located in terms of their difficulty and persons can be located in terms of their ability. The model estimates the probability of answering the item correctly as a logistic function of the difference between the person's ability and the item's difficulty. This can be seen in the formula below, which shows the form of the Rasch model for dichotomous responses:

$$P\{x_{vi} = 1 \mid \beta_v, \delta_i\} = \frac{e^{(\beta_v - \delta_i)}}{1 + e^{(\beta_v - \delta_i)}}$$

Where P depicts that the model is of a probabilistic nature, $x_{vi} = 1$ means a correct response for a particular person and item combination, and β_v and δ_i are respectively the ability of person v and the difficulty of item i

The Rasch model enables the creation of a common interval scale of scores for both the item difficulties and the person abilities, and these scores are scaled in logits. The Rasch model has the property of specific objectivity, which is a statistical form of invariance whereby the ability estimates do not depend upon the specific items used and the item difficulty estimates do not depend upon

the specific sample that were assessed.[55] This is its principle advantage over other IRT models. The Rasch model is easily extended under the same core principle to items with ordered-category (polytomous) responses. The IE analysis applies the polytomous Rasch model to certain items, as explained in the next section.

Rasch models have statistics to evaluate the fit of the item responses to the model. This can be used to determine whether to keep all of the items in the analysis, and also to provide insights into how to improve the tests for the next round.
Source: this text was partly extracted from: Cueto et al. 2009.

G.4 Rasch analysis of Kiswahili BL and ML pupil test data

This section explains the steps taken in producing the estimates of pupil ability in Kiswahili that are presented in Volume I of this report. Where relevant, it also summarises the results from key diagnostic tests that were used to assess the fit of the item response data to the Rasch model. This work builds on the Rasch analysis of BL test data reported in the BL IE report (OPM 2015b, pp97-108). The Kiswahili performance band descriptor table, which describes the skills that pupils have achieved at each band-level has been reproduced from BL, with notes to explain a few modifications revealed by the ML data.

G.4.1 Overall treatment of Kiswahili items in the Rasch analysis

At BL, the Rasch analysis of item fit led to two subtests being deleted from the analysis: listening comprehension and reading syllables (OPM 2015b, p105). A similar analysis of item fit using ML data revealed similar misfit to the Rasch model and these subtests were also excluded from the ML analysis. Item fit was primarily explored using item characteristic curves (ICCs) which compare predicted item responses from the Rasch model to observed item responses—if the data fit the Rasch model (and hence satisfy its properties) then observed item responses (for each class interval) will lie on the expected ICC curve. The ICCs of the listening comprehension items showed very poor discrimination, possibly because some of the items could be answered using common sense rather than requiring deduction from the listening passage. At BL the syllables subtest was found to systematically discriminate less than the other items, it had disordered categories[56], and there was evidence that the subtest was dimensionally divergent from the other subtests. At ML, the syllables subtest was also found to be dimensionally divergent.[57]

After dropping the two subtests, the three remaining reading subtests (familiar words, non-words, and story passage) are treated as separate polytomous items, which means that there are more than two answer categories. The answer categories are all possible reading speeds up to a cut-off speed where there were very few responses above this. All responses at or above the cut-off speed are included in one answer category. For example, on the familiar words subtest, the answer categories range from one word per minute to a top category of 46 words per minute or above.

---

[55] The Rasch model encompasses a fundamental criterion of measurement, that of invariance (specific objectivity). This requirement is independent of any particular dataset. In the case of pupil test data, the criterion of invariance means that comparison between the measures of pupil ability is independent of the set of test items used, and comparison between measures of item difficulty are independent of which pupils are used.

[56] The term 'disordered categories' means that the ordinal numbering of categories does not correspond with their substantive meaning. In this case, it meant that some slower syllable reading speeds were found higher on the scale than some faster reading speeds.

[57] Dimensionality was assessed by looking at the principle components (PC) analysis of residuals.

For the remaining subtests, each is treated as a testlet in the analysis to account for the dependence between them.58 In the analysis, testlets are treated as polytomous items with thresholds. The number of answer categories for each testlet equals the number of questions in each subtest. Answer categories are of the form 'x correct out of y questions in total'. There are 5 reading comprehension, 13 writing-spelling, and 8 writing-punctuation questions. So, for example, for reading comprehension, answer categories are 1 out of 5, 2 out of 5, 3 out of 5, 4 out of 5 and 5 out 5.

G.4.2 Steps taken in estimating Kiswahili item difficulty

This subsection briefly explains the treatment of BL item response data that was used to estimate item difficulty (i.e. the location of items on the common scale) at BL. In theory, the ML item response data should reveal similar estimates of item locations because of the criterion of invariance embedded in the Rasch model. Hence the second step is to compare estimated item locations at BL and ML. The final step reports on diagnostic tests used to reveal how well the ML item response data fits the Rasch model (when items have been anchored to the BL item locations).

1. **Recap assumptions about the treatment of non-response in the BL dataset used to estimate item difficulties**. When pupils did not respond to an item, this was treated as an incorrect response for some questions and as missing data for other questions. It is not necessary to have every pupil answer every question to estimate item difficulty accurately (because of the specific objectivity property of the Rasch model), and so where it was more difficult to determine whether pupils who did not respond to questions were in reality unlikely to know the answers, the data was coded as missing:

- Reading speed subtests: non-response is not relevant to the answer categories which simply require the number of words read correctly.

- Reading comprehension: non-response is treated as incorrect. There are two cases of non-response: the first is where the enumerator asks the pupil a question based on the passage which the pupil has just read and the pupil does not give an answer; the second is where the pupil is not asked a particular question by the enumerator because he/she did not read at sufficient speed to reach the part of the passage relevant to the question.

- Writing: in the first sentence, non-response is treated as incorrect, while in the second sentence, non-response is treated as incorrect unless all responses are non-responses; in the latter case these are treated as missing, and the pupil's response for the entire testlet treated as missing. The second sentence was automatically skipped if the pupil failed to write any word correctly in the first sentence. (Note that the treatment of missing data here is different to the treatment described in Section G.2 above for the traditional test analysis.)

Compare item locations for independent analyses of BL and ML data. The table below shows that item locations from independent Rasch analyses of the BL and ML Kiswahili item responses, are fairly similar, as expected, for all of the subtests except punctuation where the difference is 0.4 logits. This means that the punctuation subtest has become considerably easier for pupils of the same overall estimated ability across the two time points. It is difficult to know why this might have happened. One possibility is that pupils are more used to writing sentences at ML because of the new 3Rs curriculum which has writing as a subject, and so the test format is more familiar making it easier for pupils to demonstrate their skills.59

---

58 After combining the reading comprehension, spelling and punctuation items into three testlets, the residual correlations between items were acceptable.

59 This could help to explain why the estimated locations for the first two thresholds in the punctuation testlet are so much lower in the ML analysis compared to the BL, but this does not explain why the top threshold is considerably higher

Annex table 11: **Comparison of estimated Kiswahili item locations from independent BL and ML Rasch analyses**

Item	Location BL (logits)	Location ML (logits)	Difference (logits)
Reading familiar words	-0.11	-0.05	0.07
Reading non-words	0.03	0.18	0.15
Reading passage	-0.07	-0.01	0.06
Reading comprehension	0.63	0.72	0.09
Spelling	-0.57	-0.50	0.07
Punctuation	0.09	-0.35	-0.44
Source: BL and ML IE Kiswahili pupil test data.			

2. **Use BL item locations to anchor items (except for punctuation) in the Rasch analysis of ML Kiswahili item responses and then assess item fit.** The item locations in the Rasch analysis of ML test response data were constrained ('anchored') to the BL locations shown in the table above, apart from the punctuation item. Item fit was then examined primarily using ICCs. The ICCs show good fit for the three reading speed tests in particular, with observed values for all class intervals either lying on, or very close to, the ICC curve (which shows the values predicted by the Rasch model). The figure below is the ICC for the oral reading passage subtest revealing that this item fits the Rasch model well (the corresponding ICC for the BL data is in OPM 2015b, p106).

Annex figure 1: **ICC for oral reading passage subtest, ML**

ST008 SubTest 8 [ORPco]   Locn = -0.065   Spread = 0.014   FitRes = -9.659   ChiSq[Pr] = 0.113   SampleN = 2,556   Slope 34.39
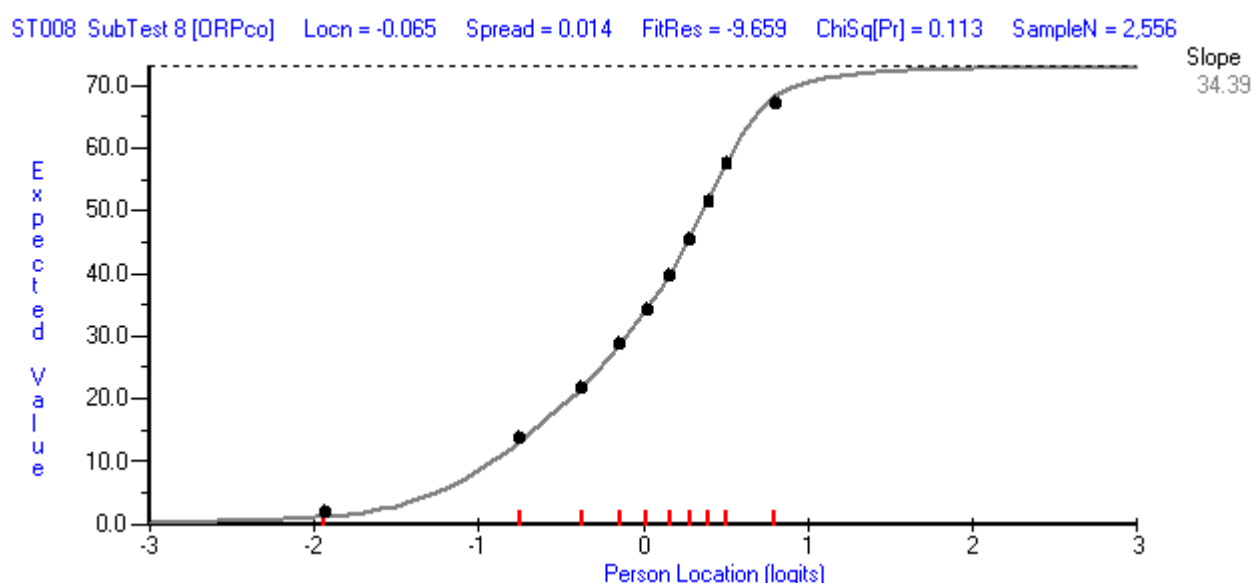
Source: ML IE survey, pupil Kiswahili test.

The worst fitting item is the reading comprehension testlet. Its ICC below shows that this item is discriminating higher than the average discrimination across all items. This was also the case with this item at BL. On balance, over-discrimination is less of a problem than under-discrimination and the misfit is not extreme, and so this item was retained.

at ML than BL. There may be some differences in the way this subtest was marked between rounds, and particular care will be taken with this at endline.

Annex figure 2:   ICC for reading comprehension subtest, ML
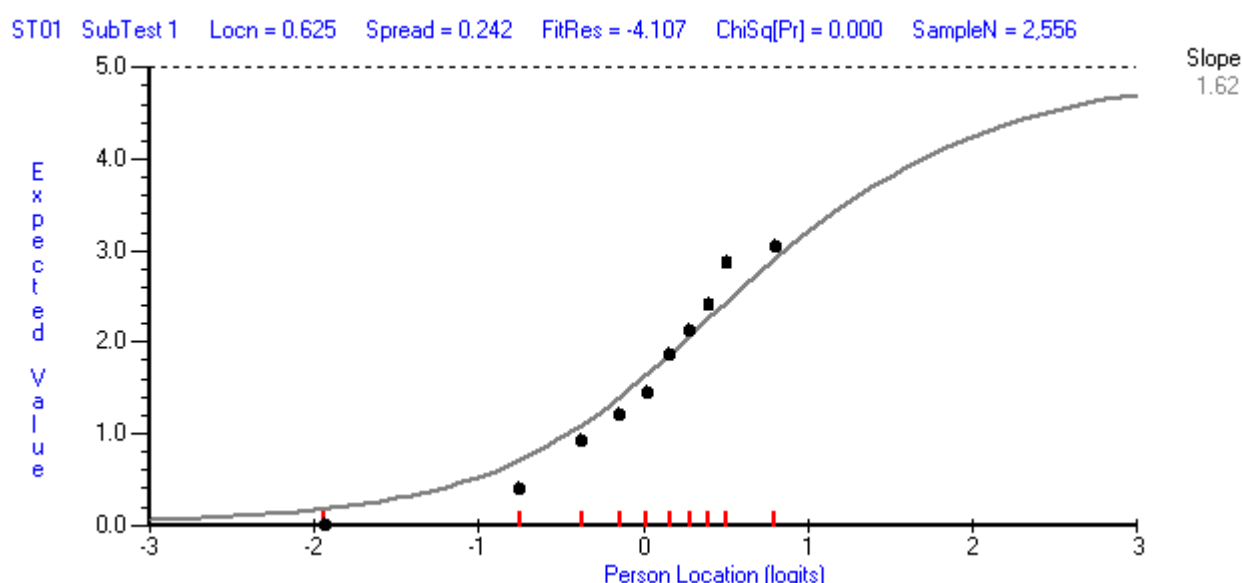
Source: ML IE survey, pupil Kiswahili test.

All of the subtests have ordered thresholds, except for non-words where at higher categories, a faster reading speed did not necessarily correspond with higher level of ability as assessed across all items. However, this issue is confined to the very top categories (above 30 words per minute) where there are far fewer observations and so the results are less reliable. This was also observed with the BL data, and is not considered serious enough to warrant deleting this item.

G.4.3   Steps taken in estimating person abilities in Kiswahili

This subsection first explains why the test data used to estimate person abilities requires different assumptions about non-response to those used above to estimate item difficulty. It then explains the steps taken to estimate person abilities (pupil Kiswahili Rasch scores) reported in Volume I, and reports on the key diagnostic tests used to examine person fit to the Rasch model. This section (Step 1) also explains why the BL pupil Kiswahili Rasch scores had to be revised slightly (although as noted in Volume I Chapter 2, the share of BL pupils falling into the different Kiswahili performance bands is almost identical to the estimates reported in the BL IE report).

1. **Make appropriate assumptions about the treatment of non-response in the pupil test data.** If non-response is treated as missing in some of the subtests (as was assumed in Step 1 above for the estimation of item difficulty), and then used to estimate person Rasch scores, it can advantage/disadvantage persons who were administered or attempted less items, and generally leads to different estimates for pupils who achieved the same overall score if they were administered and/or attempted different numbers of items. In this case, not treating missing as incorrect particularly caused some incoherence in the estimates of person ability at the lower end of the ability range, e.g., students who did not get any items correct were obtaining different estimates based on the number of items they were administered or attempted. In the case of the Kiswahili test, only the writing subtest has missing data. In order to estimate person abilities as accurately as possible, similar to the discussion in Section G.2., it was deemed reasonable to assume that pupils with missing responses were highly unlikely to have been able to write the second sentence if they were unable to write the first sentence at all. [60] So in the analysis which follows, all non-response is treated as incorrect in the estimation of person abilities. This scoring strategy was not initially applied for the BL ability estimates and

---

[60] The writing subtest consists of two sentences. The second sentence was designed to be of a similar standard to the first.

so the BL person estimates have been revised with missing treated as incorrect for the writing subtest.

2. **Re-estimate BL person abilities using BL data treated as in Step 1.**[61] This is the Rasch analysis that produces the revised BL pupil ability estimates presented in Volume I (Chapter 2). Here is a summary of the results from some of the diagnostic tests to assess fit with the Rasch model:

   Test score reliability: The person separation index (PSI, which is Rasch's equivalent of Cronbach's alpha used in traditional test analysis) is high at 0.86 demonstrating good internal consistency reliability for the test.

   Test targeting: The average difficulty of the items (constrained to be 0) was very difficult relative to the average pupil ability estimate (unweighted mean = -1.357, standard deviation 1.3961). The figure below shows the non-normal distribution of pupil ability estimates with a very large-floor effect whereby a substantial proportion of students did not answer any questions correctly.

   Person fit: the mean person fit residual is -0.160, which is fairly close to the expected value of 0, which suggests that the misfit to the Rasch model is not extreme.

Annex figure 3:   **Kiswahili person-Item distribution at BL (revised)**
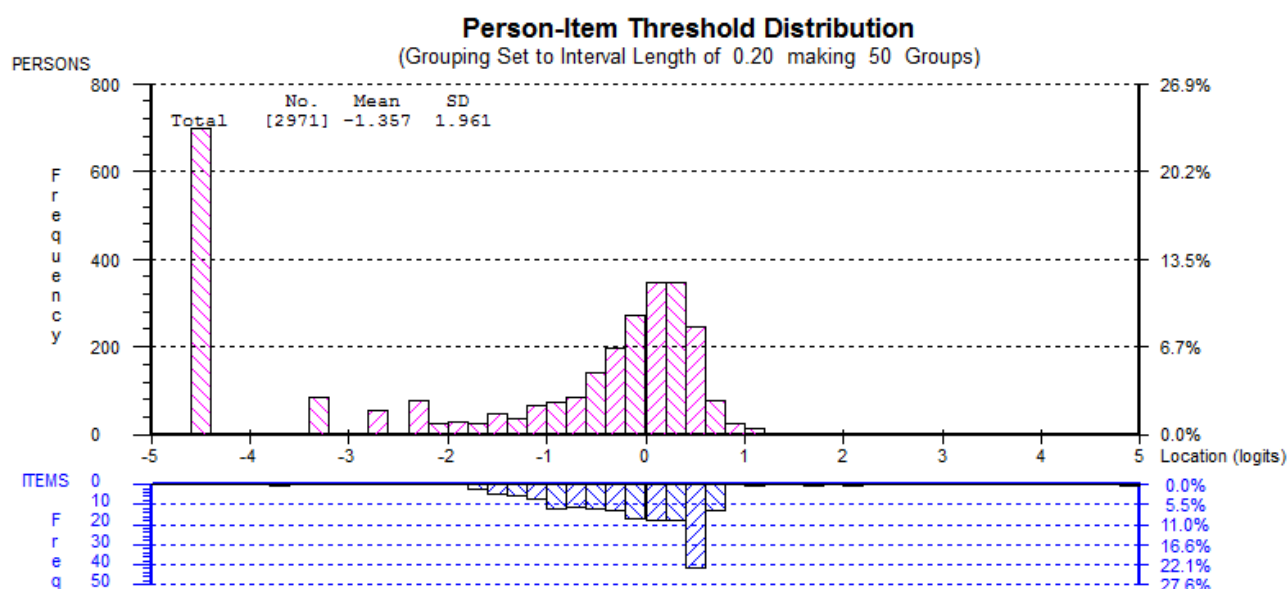
Source: BL IE survey, pupil Kiswahili test.

The original BL Kiswahili person estimates (reported in OPM 2015b, p105) had a mean person estimate of -0.912 and a standard deviation of 1.322, so, as would be expected from the treatment of the missing data, the revised analysis yields a lower mean person estimate. The correlation between the original and revised person estimates is 0.99.

3. **Estimate ML person abilities using the ML data treated as in Step 1, and use BL item locations to anchor items (except for punctuation).** This is the Rasch analysis that produces the ML pupil ability estimates presented in Volume I (Chapter 2). Here is a summary of the results from the diagnostic tests to assess fit with the Rasch model:

---

[61] The BL item locations estimated using the process outlined in Section G.4.2, were used to anchor all of the items in the revised BL analysis.

Test score reliability: The person separation index (PSI, which is Rasch's equivalent of Cronbach's alpha used in traditional test analysis) is high at 0.91 demonstrating excellent internal consistency reliability for the test.

Test targeting: The average difficulty of the items (constrained to be 0) was quite difficult relative to the average pupil ability estimate (unweighted mean = -0.587, standard deviation 1.369). The figure below shows the non-normal distribution of pupil ability estimates with a large-floor effect.

Person fit: the mean person fit residual is -0.320, which is fairly close to the expected value of 0, which suggests that the misfit to the Rasch model is not extreme.

Annex figure 4: **Kiswahili person-item distribution at ML**
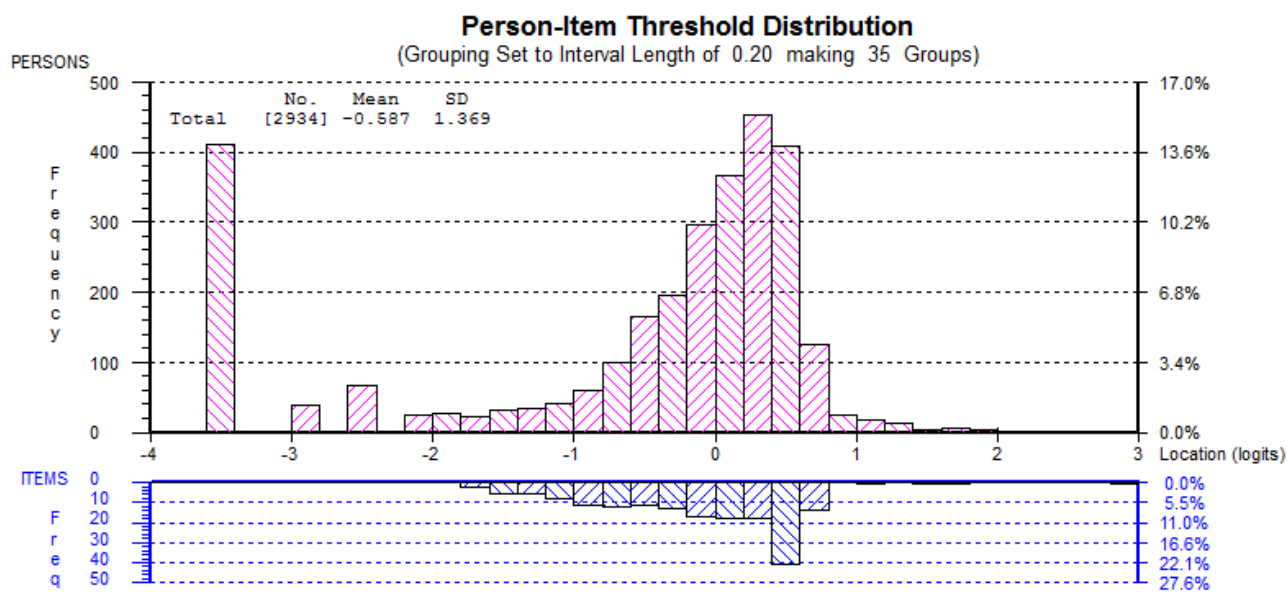
Source: ML IE survey, pupil Kiswahili test.

G.4.4 Kiswahili performance band descriptors

The description of the skills required to achieve at each of the five Kiswahili curriculum-linked performance bands has not changed since BL, and so the table below from the BL report (OPM 2015b, p102) is still applicable.

Annex table 12:   Kiswahili performance band descriptors

Score range	Items	Competency descriptor
Band 0 Below emerging skills at std 1 level		
< -1.61 logits	None	Not applicable
Band 1E Emerging skills at std 1 curriculum level: pupils have achieved at least some of the skills below		
Between -1.61 and -0.76 logits	FW: 1 to 9	Read familiar words at a speed of between 1 word and 9 words per minute
	NW: 1 to 5	Read non-words at a speed of between 1 word and 5 words per minute
	ORP: 1 to 13	*Read a simple story at a speed of between 1 and 13 words per minute*
	WSSp: 1 to 5	Spell between 1 and 5 words correctly out of 13. The spelling test included five simple short words of up to 4 letters (na,la,je, lina, letu).
	WSPu: 1 to 3	Partly punctuate sentences correctly, by getting between 1 and 3 punctuation requirements out of 8 correct. The punctuation requirements included writing text from left to right and using spacing between words.[1]
Band 1A Achieving skills at std 1 curriculum level: pupils have achieved all band 1E skills and at least some of the skills below		
Between -0.76 and -0.08 logits	FW: 10 to 20	Read familiar words at a speed of between 10 words and 20 words per minute
	NW: 6 to 13	Read non-words at a speed of between 6 words and 13 words per minute
	ORP: 14 to 30	*Read a simple story at a speed of between 14 and 30 words per minute*
	WSSp: 6 to 10	Spell between 6 and 10 words correctly out of 13, including very familiar words (shamba, shule), and simple longer words (kuvutia, darasa)
	WSPu: 4 to 5	Partly punctuate sentences correctly, by getting between 4 and 5 punctuation requirements out of 8 correct. The punctuation requirements included the use of capital letters at the start of a sentence.
Band 2E Emerging skills at std 2 curriculum level: pupils have achieved all band 1E and band 1A skills and at least some of the skills below		
Between -0.08 and 0.37 logits	FW: 21 to 30	Read familiar words at a speed of between 21 words and 30 words per minute
	NW: 14 to 21	Read non-words at a speed of between 14 and 21 words per minute
	ORP: 31 to 49	*Read a simple story at a speed of between 31 and 49 words per minute*
	RC: 1 to 2	*Answer 1 to 2 out of 5 simple reading comprehension questions correctly based on a reading a short passage, including 2 fact-based qns.*
	WSSp: 11	Spell 11 words correctly out of 13. The spelling test included simple longer words (e.g. linapendenza).
Band 2A Achieving std 2 curriculum level or above: pupils have achieved all band 1E, band 1A, and band 2E skills and at least some of the skills below		
More than 0.37 logits	FW: 31 or above	Read familiar words at a speed of 31 words or more per minute
	NW: 22 or more	Read non-words at a speed of 22 or more words per minute
	ORP: 50 or more	*Read a simple story at a speed of at least 50 words per minute*
	RC: 3 to 5	*Answer 3 to 5 out of 5 reading comprehension questions correctly based on a reading a short passage. The test included deductive and inferential questions.*
	WSSp: 12 to 13	Spell 12 to 13 words correctly out of 13. The test included simple words containing r/l (karoti) and more complex words (njegere).
	WSPu: 6 to 8	Punctuate sentences correctly, by getting between 6 and 8 punctuation requirements out of 8 correct. The punctuation requirements included the use of a full stop at the end of a sentence, and the use of a question mark at the end of a sentence.

Source: OPM 2015b, p102. Note: (1) The estimated item locations for punctuation skills differ between BL and ML (these questions were systematically easier for ML pupils of the same ability as BL pupils), but these differences do not change the bands that the different levels of punctuation skills fall into, except for the first skill level (getting 1 punctuation question correct) where the ML item location (very) narrowly falls into band 0 rather than band 1E.

G.5   Rasch analysis of maths BL and ML pupil test data

This section explains the steps taken in producing the estimates of pupil ability in maths that are presented in Volume I of this report.

G.5.1 Overall treatment of maths items in the Rasch analysis

Each question on the maths test is treated as a dichotomous item, which means that there are two answer categories: correct or incorrect. At BL, one item was dropped (number discrimination, q6) because the ICC for this item showed a pattern consistent with guessing, and poor, and at times negative discrimination, i.e., lower ability students performed better than higher ability students. This pattern was also picked up in the ML data, although the misfit was less extreme, and this item was dropped from the ML data as well.

### G.5.2 Steps taken in estimating maths item difficulties

This subsection briefly explains the treatment of BL item response data that was used to estimate item difficulty (i.e. the location of items on the common scale) at BL. In theory, the ML item response data should reveal similar estimates of item locations because of the criterion of invariance embedded in the Rasch model. Hence the second step is to compare estimated item locations at BL and ML for all 59 items, as a guide as to the items which might be functioning differently at BL and ML. Further analysis then identifies a set of items that exhibit differential item functioning (DIF) by surveywave. The final step reports on diagnostic tests used to reveal how well the ML item response data fits the Rasch model (when all items, except the set that exhibit DIF, have been anchored to the BL item locations). These steps are explained in more detail below.

1. **Recap assumptions about the treatment of non-response in the BL dataset used to estimate item difficulties**. In the BL maths dataset, all non-response is treated as missing. Most non-response occurs automatically in the test because of automatic skips (explained earlier in Section G.2). Some non-response also occurs when pupils are asked a question and they do not reply in the time allocated. The rationale for leaving the non-response data as missing when estimating item locations is that it is not necessary to have every pupil answer every question to estimate item difficulty accurately because of the specific objectivity property of the Rasch model and so no assumptions were necessary regarding the reasons for the non-responses.

2. **Compare item locations for independent analyses of BL and ML data**. The figure below shows that item locations from independent Rasch analyses of the BL and ML maths item responses are fairly similar in most cases. However, there are eight items that stand out as having an absolute difference greater than 0.3 logits: four number discrimination items (ND4,5,7,8), the first sequence item (SEQ1), two addition items (ADD11, 28) and one subtraction item (SUB28). All of the ND items, and the SEQ item, have got systematically more difficult between BL and ML. One possible reason is a change in test administration of ND items.[62] It is not clear why the addition and subtraction have become systematically easier between BL and ML.

3. **Investigate DIF by surveywave (BL to ML)**. To follow up analytically on the indication from Step 2 that some items are functioning differently at BL and ML, a Rasch analysis of the combined BL and ML dataset identified nine items which showed clear patterns of uniform DIF by surveywave. An iterative approach was then taken to split these items, starting with the item with the largest statistical indicator of DIF, in order to identify items with real DIF, as opposed to artificial DIF, which is an artefact of parameter estimation when some items have real DIF.

---

[62] In the number discrimination subtest, pupils are shown two numbers and they have to say out loud the largest number, not just point to the bigger number. It is possible that pointing was allowed at BL making it systematically easier. Careful attention will be paid to the administration of these items at endline.

Overall this analysis identified seven items with real DIF (six of these were listed in Step 2): ND4,5,7,8, SEQ1, and ADD28. The other is ADD24.

Annex figure 5:   Comparison of estimated maths item locations from independent BL and ML Rasch analyses
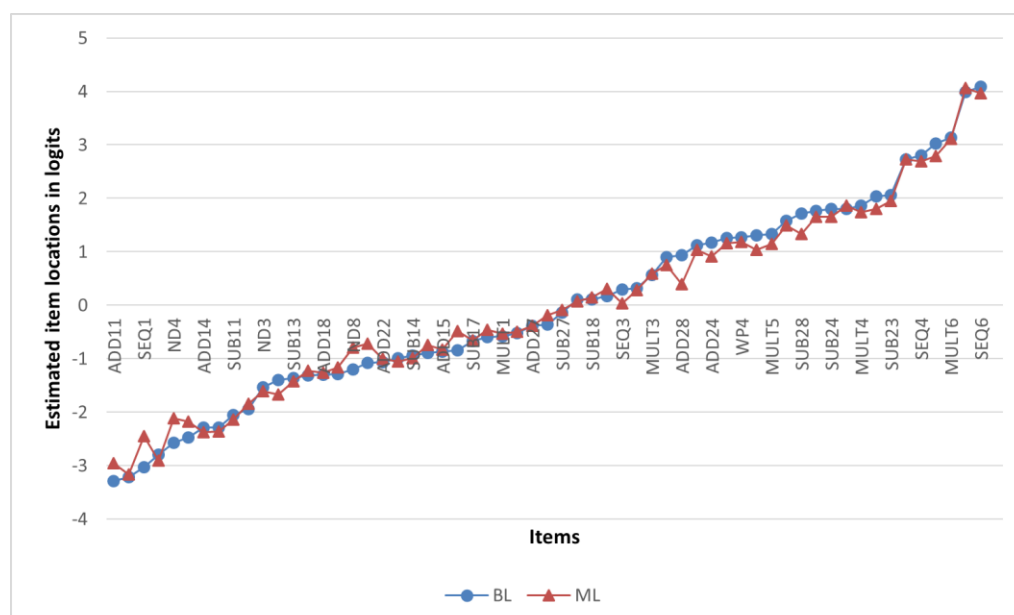
Source: BL and ML IE Kiswahili pupil test data.

4. **Use BL item locations to anchor items, except for the 7 items in Step 3 with surveywave DIF, in the Rasch analysis of ML maths item responses**. This analysis revealed a further seven items with fairly extreme patterns of misfit to the Rasch model under this anchoring approach. Hence, these seven items were also split in a combined analysis of BL and ML maths data (making 14 split items in total out of 59) and the overall item fit to the Rasch model improved.

5. **Use BL item locations to anchor items, except for the 14 items identified in Step 4, in the Rasch analysis of ML maths item responses and then assess item fit**. Item fit was examined primarily using ICCs. The ICCs show reasonable fit in most cases, with observed values for all class intervals either lying on, or not far from, the ICC curve (which shows the values predicted by the Rasch model). The figure below is the ICC for the fourth item in the word problems subtest, showing reasonable fit to the Rasch model (the corresponding ICC for the BL data is in OPM 2015b, p108). The mean item fit residual at 0.002 is close to the expected value of 0, which suggests that overall item misfit is not a serious problem.

Annex figure 6:   ICC for word problem 4, ML
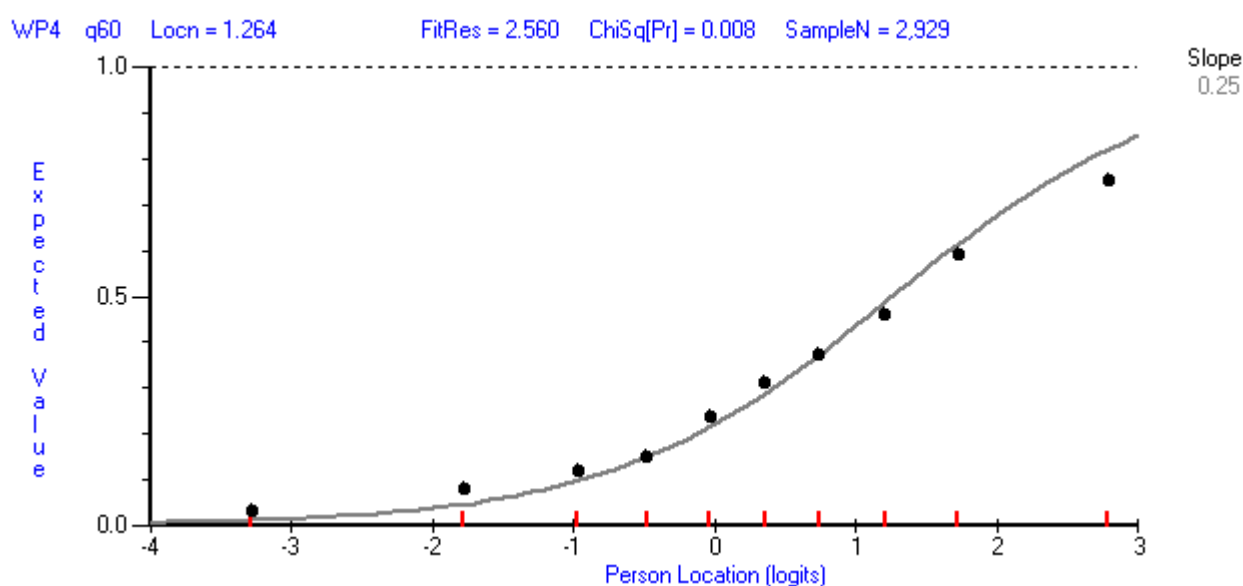
Source: IE ML survey, pupil maths test.

G.5.3 Steps taken in estimating person abilities in maths

This subsection first explains why the test data used to estimate person abilities requires different assumptions about non-response to those used above to estimate item difficulty. It then explains the steps taken to estimate person abilities (pupil Rasch maths scores) reported in Volume I, and reports on the key diagnostic tests used to examine person fit to the Rasch model. This section (Step 1) also explains why the BL pupil maths Rasch scores had to be revised (and, as noted in Volume I Chapter 2, the share of BL pupils falling into the different maths performance bands is somewhat different to that reported in the BL IE report).

1. **Make appropriate assumptions about the treatment of non-response in the pupil test data.** Similar to the rationale explained in Section G.4.3 for Kiswahili, for the purpose of estimating person abilities, it was deemed reasonable to assume that pupils with missing responses to maths items were highly unlikely to be able to answer the items correctly (because of the hierarchically difficult design of the test items as discussed in Section G.2). If the missing responses are not treated as incorrect for person ability estimates, the resulting estimates can be biased in the manner explained in Section G.4.3. So in the analysis which follows, all non-response is treated as incorrect in the estimation of person abilities. Similar to the Kiswahili BL ability estimates, the maths BL person estimates have been revised to take this into account.

2. **Re-estimate BL person abilities using BL data treated as in Step 1.[63]** This is the Rasch analysis that produces the revised BL pupil ability estimates presented in Volume I (Chapter 2). Here is a summary of the results from some of the diagnostic tests to assess fit with the Rasch model:

   Test score reliability: The person separation index (PSI, which is Rasch's equivalent of Cronbach's alpha used in traditional test analysis) is high at 0.95 demonstrating excellent internal consistency reliability for the test.

   Test targeting: The average difficulty of the items (constrained to be 0) was difficult relative to the average pupil ability estimate (unweighted mean = -0.911, standard deviation 1.910).

---

[63] The BL item locations estimated using the process outlined in Section G.5.2, were used to anchor all of the items in the revised BL analysis.

The figure below shows that the distribution of pupil ability estimates is similar to a normal bell shape but is slightly skewed to the lower ability levels.

Person fit: the mean person fit residual is -0.494, which deviates somewhat from the expected value of 0, but is not large enough to indicate serious misfit to the Rasch model.

Annex figure 7: **Maths person-Item distribution at BL (revised)**
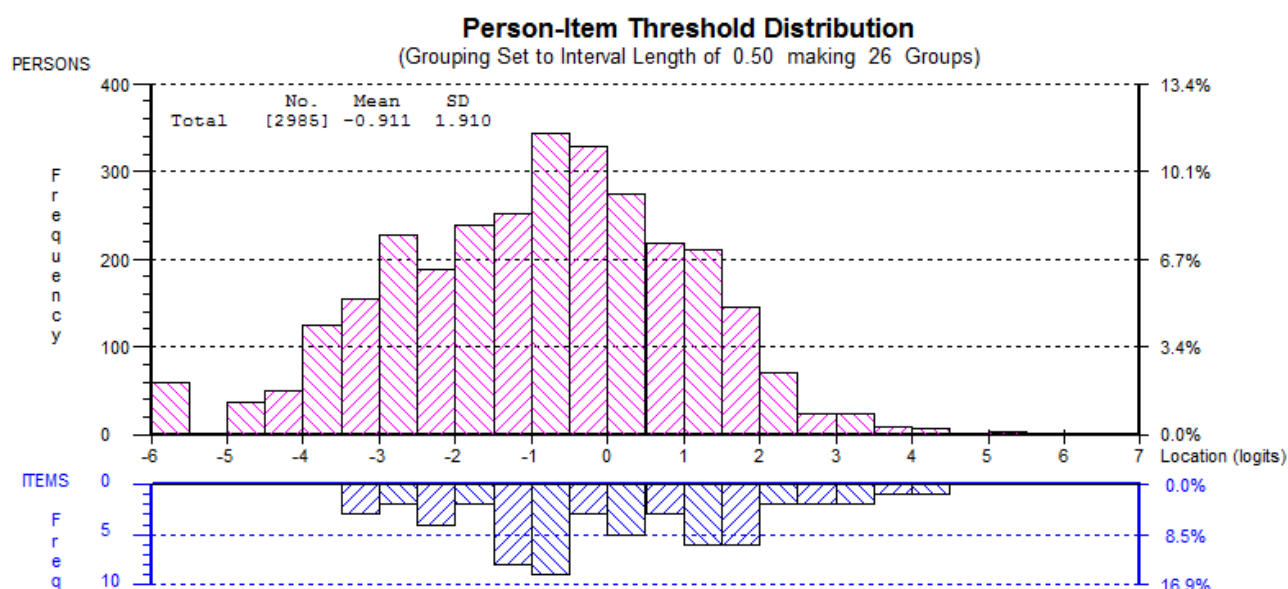
Source: BL IE survey, pupil maths test.

The original BL maths person estimates (reported in OPM 2015b, p107) had a mean person estimate of -0.541 and a standard deviation of 1.809, so, as would be expected from the treatment of the missing data, the revised analysis yields a lower mean person estimate. The correlation between the original and revised person estimates is 0.98. The effect on the share of pupils falling into the different performance bands is to increase the share falling into the lowest performance band and to reduce the shares falling into the top two performance bands. The remaining bands (band 1E and 1A) continue to account for nearly the same proportion of pupils.

3. **Estimate ML person abilities using the ML data treated as in Step 1, and use BL item locations to anchor items (except for 14 items highlighted in Section G.5.2).** This is the Rasch analysis that produces the ML pupil ability estimates presented in Volume I (Chapter 2). Here is a summary of the results from the diagnostic tests to assess fit with the Rasch model:

Test score reliability: The person separation index (PSI, which is Rasch's equivalent of Cronbach's alpha used in traditional test analysis) is high at 0.95 demonstrating excellent internal consistency reliability for the test.

Test targeting: The average difficulty of the items (constrained to be 0) was quite difficult relative to the average pupil ability estimate (unweighted mean = -0.495, standard deviation 1.909). The figure below shows that the distribution of pupil ability estimates is somewhat similar to a normal bell shape but is considerably skewed to the upper ability levels.

Person fit: the mean person fit residual is -0.418, which deviates somewhat from the expected value of 0, but is not large enough to be considered as indicative of serious person misfit.

Annex figure 8:   **Maths person-item distribution at ML**
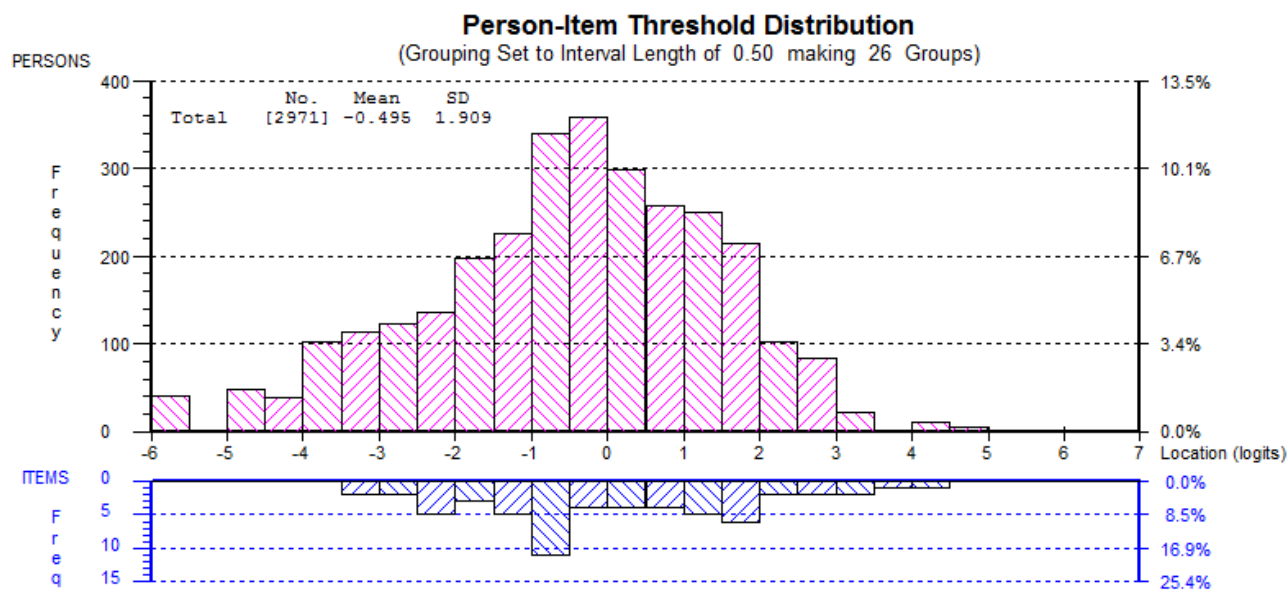
Source: ML IE survey, pupil maths test.

G.5.4 Maths performance band descriptors

The description of the skills required to achieve at each of the five maths curriculum-linked performance bands has not changed since BL, and so the table below from the BL report (OPM 2015b, p104) is still applicable. This means that although the estimated item locations for 14 out of 59 maths items differs between BL and ML (see Section G.5.2), the differences are not large enough to shift any of these skills into a different performance band.

Annex table 13: Maths performance band descriptors

Score range	Items[1]	Competency descriptor
BAND 0 Below emerging skills at std 1 level		
<-3.29 logits	None	Not applicable
BAND 1E Emerging skills at std 1 curriculum level: pupils have achieved at least some of the skills below		
Between -3.29 and -1.40 logits	ND: 1,2,3,4	Compare two whole numbers containing one or two digits, and identify which is larger
	ADD1: 1, 2,3,4	Add whole numbers containing one digit to get a total not exceeding 10
	SUB1: 1	Subtract whole numbers with values less than five
	SEQ: 1,2	Fill in missing numbers in a sequence of whole numbers containing one or two digits (less than 20) going up in steps of one
BAND 1A Achieving skills at std 1 curriculum level: pupils have achieved band 1E skills and at least some of the skills below		
Between -1.40 and -0.11 logits	ND: 5,7,8	Compare two whole numbers containing three digits, and identify which is larger (first digit is identical in both numbers, so essentially it is a comparison of two digit numbers)
	ADD1: 5,6,7,8	Add whole numbers containing one and two digits to get a total between 10 and 20
	ADD2: 1,2,7	Add whole numbers containing one, two digits and three digits to get a total between 20 and 999 (no carrying needed)
	SUB1: 2, 3, 4, 6, 7	Subtract whole numbers containing one or two digits (less than 20) (no borrowing required)
	SUB2: 1,5,7	Subtract whole numbers containing two or three digits (no borrowing needed)
	WP: 1	Solve real-life problems given in words using addition of one digit numbers to get a total not exceeding 10
	MULT: 1,2	Multiply one digit numbers with value less than five (from the 2, 3 and 4 times tables)
BAND 2E Emerging skills at std 2 curriculum level: pupils have achieved band 1E and band 1A skills and at least some of the skills below		
Between -0.11 and 2.04 logits	ADD2: 3, 4, 5, 6, 8	Add whole numbers containing two digits and three digits to get a total between 20 and 999 (carrying needed)
	SUB1: 8, 5	Subtract whole numbers containing one or two digits (less than 20) (borrowing required)
	SUB2: 2, 4, 6 ,8	Subtract whole numbers containing one, two or three digits (borrowing required)
	SEQ: 3, 5	Fill in missing numbers in a sequence of whole numbers containing two digits going up in steps of 10 Fill in missing numbers in a sequence of whole numbers containing three digits going up in steps of one
	WP: 2, 3, 4	Solve real-life problems given in words using: (i) subtraction of one digit numbers to get a total not exceeding 10; (ii) multiplication of one digit numbers to get a total not exceeding 20
	MULT: 3, 4, 5	Multiply whole numbers to get a product not exceeding 72
BAND 2A Achieving std 2 curriculum level or above: pupils have achieved band 1E, band 1A, band 2E and at least some of the skills below		
More than 2.04 logits	SUB2: 3	Subtract whole numbers containing one, two or three digits (borrowing required)
	SEQ: 4, 6, 7, 8	Fill in missing numbers in a sequence of whole numbers containing one, two or three digits: (i) going <u>down</u> in steps of two or steps of 10; (ii) going up in steps of two and five.
	MULT: 6 ,7, 8	Multiply whole numbers to get a product not exceeding 72 (including 8,9 and 12 times tables)

Source: OPM 2015b, p104. Notes (1): the items highlighted in small grey boxes (e.g. SUB1:7) are exceptions to the description given on the adjacent line.(2) Although the estimated item locations for 14 of the 59 items differ between BL and ML, these differences do not change the bands that these items fall into.

Annex H   Impact estimation supplementary information

H.1   Selection of covariates in the first stage

This annex lists linear variables that were included in the first stage selection process. Whilst the selection process was applied to each outcome indicator separately, the process started with a common initial set of linear terms for each of the two broad units of observation, i.e. the pupils and the teachers.

The sample of Standard 3 pupils who were assessed at baseline and midline form the basis of impact estimation on pupil learning. The sample constitutes a repeated cross-section of pupils in treatment schools and control schools who were interviewed at baseline and midline. Therefore, although the same treatment and control schools were visited at baseline and midline, the impact estimation at the pupil-level is not based on a longitudinal study of the same pupils over time. The evaluation setting is similar for teachers on the roster and interviewed teachers, who form the basis of impact estimation on absenteeism and school leadership management respectively. Whilst some of the teachers constitute a panel, the sample also includes teachers new to the midline and therefore, they cannot be treated as a longitudinal sample.

This has important implications for the first stage selection process of PSM: as highlighted in Chapter 6 (Section 6.2), the propensity score is valid only when it is calculated using variables unaffected by the intervention.

Although baseline data is available in the case of EQUIP-T, as mentioned earlier, the baseline and midline data are not derived from a panel of pupils (and teachers). Consequently, the range of variables that can be included in the selection model is limited to those that are unaffected by the intervention at both baseline and midline as already discussed. Since a host of intermediate variables are arguably affected by the programme, the list of covariates is reduced to static background characteristics. Annex table 5 outlines the initial set of static covariates used in building the selection model. Due to theoretical considerations, the variables respondent age and respondent gender were locked into the selection model as the literature on the subject suggests that they are key factors influencing learning outcomes.

Annex table 14:   Variables included in the first stage selection process

Analytical level	Variables
School	This mainly refers to school infrastructure-related variables such as availability of toilet, functional water source, functional electricity, library, staffroom, computers, and hygiene rating.
Head teacher	This refers to characteristics of the head teacher such as age, gender, cumulative years of experience, years of experience in the current school, professional qualifications and whether teacher is due for retirement in 2018.
Pupil (for pupil-level indicators)	This refers to pupil-level background characteristics including age, gender, presence of children in the household, presence of school going children in the household, language spoken in the household, and proxy indicators of household wealth such as ownership of radio, lantern, tables, iron and type of housing
Teacher (for teacher-level indicators)	These include gender, age, overall years of experience, number of years of experience in the current school, professional qualifications and impending retirement in 2018[1] .
Note: (1) The background characteristics (with the exception of gender) were retrospectively collected at midline for the non-interviewee baseline teachers.	

Annex I   Qualitative ML fieldwork

This annex outlines the key features of the qualitative ML fieldwork, including the selection criteria and composition of the fieldwork team, the timing and duration of fieldwork, key challenges faced by the team, and the approach taken to quality assurance in the field to ensure rigour in data collection.

The fieldwork was broadly split into two components, organised according to the types of respondents and the type of team needed. These are referred to as Team 1 – conducting the school/community visits, and Team 2 – conducting the WEC, DEO, REO, and EQUIP KIIs.

The team leaders for the two teams have prior experience of managing qualitative teams and conducting field research on education programmes, and of education planning, budgeting and management in governments respectively. The national researchers were selected based on their experience in qualitative research and development programmes.

Both teams carried out their fieldwork in April and May 2016, with overlaps within the field to discuss emerging findings.

I.1   Team 1 – School/community visits

I.1.1   The team

The school/community visits team was made up of seven core people. This included a female team leader, five national researchers (three female, two male) and a national fieldwork manager (male). The researchers were split as the team leader and interpreter for KIIs, two facilitators and two note-takers for FGDs – giving a total of three sub-teams of two people to conduct KIIs or FGDs at the same time; and the fieldwork manager organised logistics and support for the research. Of the national researchers, two were members of the team from BL and the others were selected based on their previous experience with qualitative research in Tanzania and ensuring a gender balance. The fieldwork manager worked on the quantitative survey at both BL and ML. The qualitative research leader joined the training and first week of fieldwork for quality assurance.

I.1.2   Training

This team began on 18 April 2016 with one week of training, including a day of piloting in an EQUIP-T school in Dodoma. The training began with an introduction to the EQUIP-T programme, a refresher on qualitative research approach and methods, the research questions, design and tools, and ethical considerations. The following two days focused on the instruments for KIIs and FGDs, working on the translation and carrying out role plays. The logistics for fieldwork and principles for managing the data were covered. Day 4 was used as a pilot, for all team members to test their instruments in an EQUIP-T school. The final day was used for debrief and revisions to the instruments, as well as discussing challenges faced during the pilot and how to mitigate these during data collection.

I.1.3   Data collection

The team spent three weeks carrying out school visits – one week in each district – with one full day in each of three schools and two days per district to work on debriefs and transcripts. The fieldwork followed around one week after the quantitative team had visited.

Head teachers were informed of the arrival of the team ahead of schedule and were reminded of the exact dates for the research. During a school visit, the team carried out two KIIs (the HT and community leader) and five FGDs (teachers, pupils, mothers, fathers, school committee members). The support of head teachers was enlisted in mobilising interviewees and FGD participants. The selection of the participants for the FGDs and KIIs is given in

Table 5 above.

At the start of each visit day, two researchers would observe one whole lesson, looking at the teaching methods being used, teaching aids, what was in the classroom, and inclusive instruction including gender. The same two researchers ran the pupils' FGD, so they could probe and follow up on what they had seen in the lesson. Female researchers were chosen for this so it would be easier for girls to participate in discussions. The female researchers also ran the FGD with mothers, for the same reason. The whole team carried out observations throughout the day, particularly during break times, to feed into debriefs.

I.1.3.1 Challenges during data collection

There were a few challenges during data collection:

First, the sampling of fathers, mothers and school committee members was affected as not all invited respondents attended. To adjust to this, FGDs had to be re-structured into group interviews in some cases, where only three respondents were available. Similarly, some schools only had three teachers who had attended EQUIP-T INSET (when excluding head teachers), and thus a similar adjustment was made in this cases. Realising this challenge, the team leader ran a refresher training for the focus group facilitators in order to emphasis the different focus of a group interview vis-à-vis a FGD.

Second, due to the vast scope of the evaluation matrix (as a consequence of the need to gather data around all four components of EQUIP-T), KIIs and FGDs would be lengthy. This ran the risk of respondent-fatigue. The team would mitigate this through offering respondents to take breaks, as well as beverages. Beverages were not offered as an incentive to take part, but rather during the interview with the purpose of mitigating respondent-fatigue. Moreover, the lengthy fieldwork in combination with long instruments ran the risk of researcher-fatigue. The daily debriefs were used to mitigate this by incorporating refresher training, motivating researchers by including them in wider analysis through regional debriefs, and by ensuring that researchers had evenings off where no de-briefs were scheduled, and transcription postponed. The importance of mitigating researcher-fatigue should not be underestimated, as it may impact on researchers' motivation to probe effectively as findings emerge during data collection, thus missing out on essential analysis.

I.1.3.2 Quality Assurance processes

Quality assurance was carried out in the field through a number of mechanisms. First, the team leader was available in the field at all times to be consulted, and to make decisions and technical adjustments as needed whilst ensuring consistency in the conceptual framework and design.

Second, debriefs were conducted following each school visit. In debriefs, the team would discuss the day's key findings, link findings to the evaluation matrix and compare observational notes to help situate analysis. Debriefing sessions were also used to address logistical and technical queries that may have arisen during the day and challenge researcher perceptions and biases. Additionally, the team held briefings each morning before visiting schools in order to give technical pointers for the day ahead and identify questions needing further probing and exploration.

In order to ensure quality transcription, a separate training session was held focusing on the importance of *per verbatim* transcription, standardisation of translation, and to give researchers tips and tricks with regards to time management and speed of transcription whilst maintaining accuracy. Additionally, any discrepancies in translation or queries regarding transcription were discussed during the daily debriefs. The team lead was also available at all times to help the team with queries throughout fieldwork. Transcriptions were done by the researchers in the days before visiting the next school. This was to allow the team lead to quality assure samples of transcripts in

the field, ensuring rigour in data collection. This also allowed the team lead to give guidance (particularly prior to going to the next district), ensuring reflexivity in the field. Upon returning from the field, the team lead reviewed all transcripts, compared them to field notes and made use of back-translation in cases of discrepancy. All transcription was done *per verbatim* in order to mitigate bias at the stage of data capturing.

I.2 Team 2 – Senior manager interviews

I.2.1 The team

The programme, regional, district, and ward management interviews were conducted by a smaller team consisting of a team leader and a note-taker/interpreter.

I.2.2 Training

A brief training was carried out for the interpreter on the background to the programme, the research questions and objectives of the evaluation, the principles of qualitative research including ethics, and techniques for translation.

I.2.3 Data collection

Interviews with EQUIP-T programme staff at headquarters were carried out over two days by the team leader, after which the pair visited the three regions/districts over the course of ten days. The RTLs, REOs and DEOs were called in advance to confirm the date and time, and to get hold of contact numbers for the relevant WECs. All interviewees were called again the day before to confirm. WECs were asked to come to the district centre for the interview, and a quiet, private room was provided in each case. Interviews with regional level and above were conducted in English, but DEOs and WECs generally spoke in Kiswahili with translation by the interpreter. The team leader tried to keep all interviews to a maximum of two hours to prevent respondent fatigue.

After each interview, the two team members recorded key observations of the interviewee or the context of the interview. The team discussed major themes and supporting or contradicting perceptions between different respondents, and identified areas needing further probing.

Annex J   Costing study data tables

J.1   PSA spending

Annex table 15:   Spending up to June 2016 and budgets for programme support activities by EQUIP-T MA

	Spending (£m)	% of total spending	Budget (£m)	% of budget	Budget execution
C1 Improving the Performance of Teachers	5.04	47%	11.38	32%	44%
C2 Strengthening School Leadership and Management	0.81	8%	3.95	11%	21%
C3 District Planning and Management	1.89	18%	10.01	28%	19%
C4 Community Participation & Demand for Better Accountability	0.72	7%	6.42	18%	11%
C5 Strengthened Learning & Dissemination of Results	2.32	22%	4.23	12%	55%
Total	10.78	100%	36.00	100%	30%
Source: EQUIP-T MA Budget tracker					

Annex table 16:   Spending on programme support activities by EQUIP-T MA in implementation, up to June 2016

	Subcomponent	Spending GBP
Component 1	1.1 Developing a Teacher Performance Framework	11,695
	1.2 Improving the Performance of Teachers	3,952,527
	1.3 Developing a Teacher Performance Management System	2,986
	1.4 Improving Teacher Morale	157,555
	1.5 School Readiness Programme	915,101
	Total component 1	**5,039,864**
Component 2	2.1 Developing a School Quality Framework and Leadership Performance Framework	89,372
	2.2 Design and Implementation of Head Teacher Performance Management System	178,183
	2.3 Strengthening Head Teacher and WEC School Leadership and Management	541,605
	2.4 Peer support for Whole School Development	2,837
	Total component 2	**811,997**
Component 3	3.1 EQUIP-T set up, Baseline and Programming	19,940
	3.2 Strengthening District Planning and Management Capacity	340,229
	3.3 Support Districts to prepare to management EQUIP-T programmes from 2016 and plan for replication and scale-up	3,087
	3.4 Support Districts to manage, co-ordinate and monitor special activities/ grants	1,527,912
	Total component 3	**1,891,168**

Component 4	4.1 Establish relationships with CSO partners (HakiElimu, Twaweza/ UWEZO etc.)	76,349
	4.2 Conduct Community Level Consultations, Introduction of EQUIP-T programme	-
	4.3 Core Activity 1: Improve communications mechanism for communities	42,398
	4.4 Core Activity 2: Community engagement to education planning	464,085
	4.5 Core Activity 3: Build capacity of WECs to train SCs/ build capacity for effective operations of the school committee	131,701
	4.6 Support to link community education plan objectives into School Development Plan (collaboration with Leadership Unit)	131
	4.7 Core Activity 4: PTA formation	2,758
	4.8 Core Activity 5: Development of School IGA	1,214
	4.9 Support PTA to establish school clubs/ student parliaments/ interest groups	1,803
	4.10 Further develop Transparency, Accountability and representation mechanisms and projects	-
	4.11 Continue advocacy and communication campaigns	-
	4.12 Roll-out community score cards in line with Core activity 1 and in collaboration with Planning, Leadership and Equity Units.	-
	4.13 Continue capacity building programmes and maintenance (sharing, exchange visits, refresher training, etc.)	-
	Total component 4	**720,439**
Component 5	5.1 EMIS	1,792,517
	EMIS Contingency money 14C10-115	147
	5.2 Programme Monitoring & Evaluation	468,789
	5.3 Programme Communications	59,785
	5.4 Evaluation Auditing	14
	5.2 staff capacity strengthened to provide both real time and periodic system information	-
	9.4.4012 - Stationery, Printing, Communications etc	1,706
	Total component 5	**2,322,959**
	Total	**10,786,426**

Source: EQUIP-T MA Budget tracker

Annex table 17: Unit costs of PSA activities paid for by the MA, up to June 2016

Component	Unit costs	GBP	Notes
Component 1 – Teachers	Spending on improving teachers, per teacher trained on curriculum	317	A
	Spending on improving teachers, per child in primary school	1.69	A
	Spending on improving teachers, per school	925	A
	Spending on SRP per child enrolled in SRP	18	B
Component 2 – SLM	Strengthening school leadership and management - per trainee	88	C
	Strengthening school leadership and management - per school	182	

Component 3 – DRM	Strengthening district planning and management, per district	39,399	D
Component 4 – Communities	Strengthening community participation, per school	162	E
Component 5 – Learning	Strengthened learning and information, per school	521	F

Notes: A is based on all C1 spending minus SRP. B uses spending on component 1.5. C uses all C2 spending, divided by SLM2 trainees. D uses all C3 spending, divided by 48 districts. E uses all C4 spending. F uses all C5 spending. Where units are per school, the number of schools was 4,460, taken from EQUIP-T monitoring data.

J.2   Sub-component 1.2

Annex table 18:   Spending by EQUIP-T MA under sub-component 1.2, broken down by detailed codes and by quarter

Code	Description	Spending (GBP)										% of 1.2	Entries
		2014 Q2	2014 Q3	2014 Q4	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	2016 Q2	Total		
1.2.1	**Conduct M+E baseline study on teacher performance**												
1.2.1.4	Data collection, compilation and management							223			223	0%	2
1.2.2	**Review ongoing INSET initiatives and establish synergies**												
1.2.2.2	Presentation/consultation (individuals and consultative groups)		621			1,159					1,779	0%	2
1.2.2.3	Workshop - Presentation/consultation	6,253	3,474			288	4,137	687		1,485	16,325	1%	8
1.2.2.4	Draft Paper - consultation findings							938	36		974	0%	4
1.2.2.6	Consult model	9				349		1,943	1,895		4,196	0%	6
1.2.3	**Commission universities/TTC as service providers**												
1.2.3.5	Contract award - Universities				47,487		28,182	5,987	16,177	1,621	99,453	3%	10
1.2.3.6	Contract award - TTC	3,468		149,529	805,395	35,436	28,269	25,952	25,098	12,342	1,085,489	34%	106
1.2.4	**Develop Kiswahili literacy INSET module**												
1.2.4.3	Feedback received and first draft produced		253					52			305	0%	3
1.2.4.5	Draft modules and technologies trialled				621	738	2,353	596			4,309	0%	12
1.2.4.6	Module adapted, final draft approved								121		121	0%	1
1.2.4.7	Module printed and distributed		10,305	10,192		854		166	122		21,639	1%	7
1.2.4.8	Procure Teaching Aid Toolkit			632	53,908		215,632			28	270,200	8%	4
1.2.4.9	Literacy materials procured and distributed				15,409	82,983		69,728	67,455	19,368	254,943	8%	15
1.2.5	**Train District INSET Teams**												
1.2.5.1	Prepare facilitation manual and training plan with University partners									76,887	76,887	2%	3
1.2.5.2	Print training and orientation materials					422			106	630	1,158	0%	3
1.2.5.3	5-day training of District INSET Teams		2,666		9,608	2,647	3,145	6,800		410	25,275	1%	39
1.2.6	**Conduct Regional introduction workshops on school-based INSET system**												
1.2.6.2	Orientate REOs, DEOs, district officers and inspectors on school-based INSET	1,091	4,354	16,256						22	21,724	1%	36

Code	Activity										Total	%	No.
1.2.7	**Introduction/refresher training for INSET coordinators, HTs**									12,919			
1.2.7	Introduction/refresher training - Other									12,919	12,919	0%	7
1.2.7.2	Implement intro/refresher training	6,267		23,396	3,103		23,644	5,586	466	3,790	66,252	2%	30
1.2.8	**Introduction and implementation of school-based INSET for St. 1-3**												
1.2.8.2	Ward-cluster launch workshops for St. 1-3 teachers	967	22,280	43,168	8,547	164,305	312,749	34,985	96,826	57,271	741,097	23%	543
1.2.8.3	School-based INSET for St. 1-3 teachers			91	39,288	363			502		40,244	1%	14
1.2.10	**Develop early grade maths module and Inclusive/gender responsive modules**												
1.2.10	Develop early grade maths module and Inclusive/gender responsive modules	254									254	0%	1
1.2.10.	Develop early grade maths module - other			1,138			4,466				5,604	0%	2
1.2.10.1	Produce outline of maths modules for initial consultation						513				513	0%	1
1.2.10.6	Modules printed and distributed		858		73,498	60,452	93,970		135,506	4,335	368,618	11%	14
1.2.12	**Train INSET coordinators/ subject specialists and launch St. 4-7 INSET**						628						
1.2.12.	Train INSET coordinators - other						4,382	210			4,592	0%	4
1.2.12	Train INSET coordinators/subject specialists and launch St. 4-7 INSET						628				628	0%	1
1.2.12.4	Ward cluster launch workshops for St. 4-7 maths/science teachers			9,428		1,655	33,474		9,550	49,548	103,656	3%	60
1.2.13	**Innovation fairs for teachers in Jan 2015**												
1.2.13	Innovation fairs for teachers in Jan 2015									4,528	4,528	0%	2
1.2.13.3	Produce and distribute booklet (print/digital format) of most innovative teacher activities				2,339		2,767				5,105	0%	2
1.2.14	**Promote development of teacher-initiated INSET modules/classroom research**												
1.2.14.1	Produce guidance on the development of modules. Produce guidance with case studies of classroom research									308	308	0%	1
1.2	**Other**								316				
1.2	Other								316		316	0%	3
Total	Total	18,309	44,811	253,831	1,059,203	350,492	760,095	153,643	354,580	258,531	3,239,634	100%	946

Source: EQUIP-T MA 'Quickbooks' database. Notes: Costs in GBP calculated using the exchange rate on 30 June 2016 from xe.com

Annex table 19:   **Spending by EQUIP-T MA under sub-component 1.2, on the largest codes**

Code	Amount GBP	%
1.2.3 Commission universities/ TTC as service providers	1,184,942	37%
1.2.8 Introduction and implementation of school-based INSET for St. 1-3	781,341	24%
1.2.4 Develop Kiswahili literacy INSET module	551,518	17%
1.2.10 Develop early grade maths module and Inclusive/gender responsive modules	374,475	12%
Other*	347,650	11%

Source: EQUIP-T MA 'Quickbooks' database.

Notes: Costs in GBP calculated using the exchange rate on 30 June 2016 from xe.com.

*Other includes categories each less than 3.3% of spending: 1.2.1 Conduct M+E baseline study on teacher performance; 1.2.2 Review ongoing INSET initiatives and establish synergies; 1.2.5 Train District INSET Teams; 1.2.6 Conduct Regional introduction workshops on school-based INSET system; 1.2.7 Introduction/refresher training for INSET coordinators, HTs, (Facilitating/mentoring skills); 1.2.10 Develop early grade maths module and Inclusive/gender responsive modules; 1.2.12 Train INSET coordinators/subject specialists and launch St. 4-7 INSET; 1.2.13 Innovation fairs for teachers in Jan 2015; 1.2.14 Promote development of teacher-initiated INSET modules/classroom research publications; 1.2 Other

J.4 LGA spending

Annex table 20:   Spending by LGAs by activity/Epicor code and region, up to May 2016 (GBP)

Activities	EPICOR code per activity	DODOMA REGION	KIGOMA REGION	LINDI REGION	MARA REGION	SHINYANGA REGION	SIMYU REGION	TABORA REGION	Total
3Rs (KKK) INSET*	C04C01	453,073	428,521	162,939	307,246	440,995	283,608	509,381	**2,585,764**
Community and school partnerships	C04C02	-	1,238	5,378	14,857	614	-	-	**22,087**
Education grant management and planning	C04C03	-	-	-	-	-	-	-	**-**
INSET contracting of teacher training colleges	C04C04	-	8,890	9,647	4,800	3,786	6,535	41,215	**74,873**
INSET general	C04C05	-	-	-	-	-	-	-	**-**
LGA education planning and management	C04C06	-	-	-	-	-	-	-	**-**
PTP grants	C04C07	137,311	120,053	92,363	142,432	105,449	97,673	111,708	**806,988**
School IGAs	C04C08	-	-	-	-	8,083	-	3,125	**11,208**
School leadership and management	C04C09	189,648	74,243	190,945	58,216	33,693	47,576	93,849	**688,171**
WEC grants	C04C10	24,481	61,782	37,180	84,001	62,480	61,354	28,372	**359,648**
Other costs	-	-	-	-	-	-	-	-	**-**
Total		**804,513**	**694,727**	**498,451**	**611,552**	**655,100**	**496,745**	**787,650**	**4,548,739**

Source: EQUIP-T MA consolidation reports of LGA spending up to May 2016.
Notes: Costs in GBP calculated using the exchange rate on 30 June 2016 from xe.com.
*3Rs INSET is the label given by EQUIP-T and relates to the early grade reading and maths (EGR, EGM) modules developed and cascaded by EQUIP-T.

Annex table 21:   Budget execution: proportion of budgets spent by LGAs by activity/Epicor code and region, up to May 2016 (%)

Activities	EPICOR code per activity	DODOMA REGION	KIGOMA REGION	LINDI REGION	MARA REGION	SHINYANGA REGION	SIMYU REGION	TABORA REGION	Total
3Rs (KKK) INSET*	C04C01	119.6	136.3	40.5	50.4	154.5	113.9	126.7	97.9
Community and school partnerships	C04C02	0.0	0.8	4.0	7.4	0.4	0.0	0.0	1.9
Education grant management and planning	C04C03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
INSET contracting of teacher training colleges	C04C04	0.0	3.4	3.6	1.5	1.8	3.3	18.7	4.3
INSET general	C04C05	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LGA education planning and management	C04C06	0.0	n/a	n/a	n/a	n/a	n/a	n/a	0.00
PTP grants	C04C07	107.9	100.8	101.0	97.8	99.6	100.0	77.2	97.0
School IGAs	C04C08	0.0	0.0	0.0	0.0	14.6	0.0	4.1	2.4
School leadership and management	C04C09	73.0	30.6	98.8	19.6	15.5	23.7	32.0	40.4
WEC grants	C04C10	16.1	64.5	34.2	62.6	57.5	58.8	17.2	41.4
Other costs	-	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Total		**46.7**	**47.6**	**35.1**	**30.0**	**49.7**	**40.9**	**44.6**	**41.6**

Source: EQUIP-T MA consolidation reports of LGA spending up to May 2016.

Notes: *3Rs INSET is the label given by EQUIP-T and relates to the early grade reading and maths (EGR, EGM) modules developed and cascaded by EQUIP-T.

Annex table 22: **Spending by LGAs by activity/Epicor code, up to May 2016 (GBP)**

LGA	3Rs (KKK) INSET*	Community and school partnerships	Education grant management and planning	INSET contracting of TTCs	INSET general	LGA education planning and management	PTP grants	School IGAs	School leadership and management	WEC grants	Total
	C04C01	C04C02	C04C03	C04C04	C04C05	C04C06	C04C07	C04C08	C04C09	C04C10	
DODOMA											
Bahi DC	53,896	-	-	-	-	-	13,655	-	8,972	3,041	79,565
Chamwino DC	67,845	-	-	-	-	-	22,569	-	19,548	6,841	116,803
Chemba DC	68,373	-	-	-	-	-	19,535	-	21,549	3,345	112,801
Dodoma Urban	61,893	-	-	-	-	-	17,448	-	30,273	1,762	111,377
Kondoa DC	80,552	-	-	-	-	-	22,379	-	32,759	7,931	143,622
Kongwa DC	71,760	-	-	-	-	-	19,914	-	24,866	1,560	118,100
Mpwapwa DC	48,755	-	-	-	-	-	21,810	-	51,681	-	122,247
KIGOMA											
Buligwe DC	52,096	-	-	707	-	-	16,690	-	11,501	7,697	88,690
Kakonko DC	37,056	-	-	1,500	-	-	11,190	-	6,466	5,138	61,349
Kasulu DC	84,538	-	-	586	-	-	25,414	-	15,326	14,538	140,402
Kibondo DC	55,707	-	-	1,517	-	-	15,741	-	14,600	9,554	97,120
Kigoma DC	61,812	-	-	1,034	-	-	20,104	-	7,229	5,793	95,972
Kigoma-Ujiji MC	64,882	1,238	-	897	-	-	8,535	-	4,586	7,697	87,834
Uvinza DC	72,431	-	-	2,649	-	-	22,379	-	14,535	11,366	123,360
LINDI											
Kilwa DC	36,136	-	-	2,812	-	-	19,724	-	36,143	9,993	104,809
Lindi DC	35,443	-	-	1,834	-	-	21,242	-	54,087	8,800	121,405
Lindi TC	8,886	625	-	-	-	-	5,879	-	11,852	1,414	28,657

Liwale DC	19,474	1,345	-	1,628	-	-	10,241	-	19,222	4,345	56,255
Nachingwea DC	32,117	1,640	-	2,855	-	-	19,914	-	37,793	9,514	103,833
Ruangwa DC	30,882	1,769	-	517	-	-	15,362	-	31,848	3,114	83,493
MARA											
Bunda DC	68,946	2,436	-	-	-	-	30,345	-	10,090	17,959	129,777
Butiama DC	49,512	990	-	-	-	-	16,121	-	6,143	10,476	83,241
Musuma DC	-	1,131	-	-	-	-	20,483	-	7,250	13,448	42,312
Musuma MC	9,636	4,885	-	-	-	-	7,207	-	-	3,421	25,148
Rorya DC	68,400	2,040	-	-	-	-	22,759	-	10,290	11,117	114,606
Serengeti DC	53,147	1,545	-	2,793	-	-	20,673	-	12,280	19,242	109,679
Tarime DC	50,750	1,293	-	2,007	-	-	19,724	-	11,419	4,917	90,111
Tarime TC	6,854	538	-	-	-	-	5,121	-	744	3,421	16,677
SHINYANGA											
Kahama TC	60,866	-	-	621	-	-	13,655	5,762	-	8,690	89,594
Kishapu DC	49,065	124	-	276	-	-	21,810	-	9,062	12,672	93,010
Msalala DC	76,167	-	-	759	-	-	17,259	-	6,041	9,124	109,350
Shinyanga DC	120,226	-	-	-	-	-	24,466	-	10,852	9,683	165,226
Shinyanga MC	64,777	490	-	621	-	-	9,104	-	7,738	9,848	92,577
Ushetu DC	69,894	-	-	1,510	-	-	19,155	2,321	-	12,462	105,343
SIMIYU											
Bariadi DC	40,830	-	-	1,448	-	-	13,845	-	6,798	7,386	70,307
Bariadi TC	22,386	-	-	-	-	-	7,207	-	3,750	4,310	37,653
Busega DC	46,845	-	-	1,534	-	-	16,310	-	7,476	12,828	84,994
Itilima DC	49,531	-	-	517	-	-	16,500	-	8,316	9,407	84,271
Maswa DC	64,595	-	-	1,483	-	-	22,759	-	10,879	15,393	115,109
Meatu DC	59,421	-	-	1,552	-	-	21,052	-	10,357	12,029	104,411

TABORA											
Igunga DC	73,268	-	-	1,293	-	-	25,224	-	3,379	5,559	108,724
Kaliua DC	89,126	-	-	224	-	-	18,017	-	7,771	12,488	127,627
Nzega DC	83,061	-	-	37,753	-	-	-	-	28,429	-	149,242
Nzega TC	-	-	-	-	-	-	-	-	-	-	-
Sikonge DC	53,589	-	-	-	-	-	18,017	-	3,056	1,379	76,042
Tabora MC	40,960	-	-	-	-	-	13,845	-	8,007	-	62,812
Uyui DC	108,758	-	-	-	-	-	22,000	3,125	15,097	3,393	152,373
Urambo DC	60,620	-	-	1,945	-	-	14,604	-	28,111	5,552	110,831
Total	2,585,764	22,087	-	74,873	-	-	806,988	11,208	688,171	359,648	4,548,739

Source: EQUIP-T MA consolidation reports of LGA spending up to May 2016.

Notes: Costs in GBP calculated using the exchange rate on 30 June 2016 from xe.com. *3Rs INSET is the label given by EQUIP-T and relates to the early grade reading and maths (EGR, EGM) modules developed and cascaded by EQUIP-T.

Annex table 23:   Distribution of LGAs' Epicor activities across EQUIP-T's five components

Component	Activities/codes included
Component 1	3Rs (KKK) INSET INSET contracting of teacher training colleges INSET general
Component 2	School leadership and management
Component 3	Education grant management and planning LGA education planning and management WEC grants
Component 4	Community and school partnerships PTP grants School IGAs
Component 5	None applicable

Annex table 24: Spending by LGAs by component, units of beneficiaries, and average spending or unit costs

	DODOMA	KIGOMA	LINDI	MARA	SHINYANGA	SIMIYU	TABORA	TOTAL
Spending (GBP)								
Component 1	453,073	437,412	172,586	312,046	444,781	290,143	550,596	2,660,637
Component 2	189,648	74,243	190,945	58,216	33,693	47,576	93,849	688,171
Component 3	24,481	61,782	37,180	84,001	62,480	61,354	28,372	359,648
Component 4	137,311	121,290	97,741	157,289	114,146	97,673	114,833	840,283
Component 5	-	-	-	-	-	-	-	-
Total	**804,513**	**694,727**	**498,451**	**611,552**	**655,100**	**496,745**	**787,650**	**4,548,739**
Units								
Enrolment	394,587	393,619	171,944	442,452	273,660	313,978	446,966	2,437,206
Number of schools	727	637	493	759	559	516	769	4,460
Teachers trained EGR Modules 5-8	1435	1266	1119	Not available	1130	1472	1451	7873
Number trained on SLM2	1640	1408	1124	923	1237	1148	1719	9199
Number of wards	194	141	139	177	127	145	208	1131
Number of LGAs	7	7	6	8	6	6	8	48
Unit costs (GBP)								
INSET spending per teacher trained (EGR Modules 5-8)	316	346	154	N/A	394	197	379	338
INSET spending per pupil	1.15	1.11	1.00	0.71	1.63	0.92	1.23	1.09
INSET spending per school	623	687	350	411	796	562	716	597
SLM spending per HT/AHT/WEC trained	116	53	170	63	27	41	55	74
SLM spending per school	261	117	387	77	60	92	122	154
C3 spending per LGA (Tsh millions)	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01
Average WEC grant	126	438	267	475	492	423	136	318
C4 spending per school	189	190	198	207	204	189	149	188
- of which PTP grant per school	189	188	187	188	189	189	145	181

Source: EQUIP-T MA consolidation reports of LGA spending up to May 2016, and EQUIP-T MA Summary Master Data (monitoring data) – Fact sheet March 2016, document dated July 2016. Notes: Costs in GBP calculated using the exchange rate on 30 June 2016 from xe.com.

J.5   Costs of INSET

Annex table 25:   Early Grade INSET courses run by three LGAs, deriving unit costs

Label	LGA	Month (2015-16)	Module	Delivery method	Number of days*	Number attended	Who were they*	Total cost (Tsh)	Unit cost total course (Tsh)	Unit cost per day (Tsh)*	Unit cost per day (GBP)* +
A	Tabora MC	January	EGR 5-13	Residential Municipal	4	278	St1/2 teachers, ICs	67,502,860	242,816	60,704	20.9
B		January	EGR 5-8 review	Clusters (9)	1	175	WECs, ICs and HTs	5,405,000	30,886	30,886	10.7
C		April	EGR 9-12 review	Clusters (10)	1	175	WECs, ICs and HTs	5,845,000	33,400	33,400	11.5
D		April	EGM 1-4	Residential Municipal	5	253	ICs, Key Maths teacher, QA (4 days), WECs and HTs (1 day)	54,405,000	215,040	43,008	14.8
E		May	EGM 1-2 review	Clusters (9)	1	174	WECs, ICs and HTs	6,420,000	36,897	36,897	12.7
F	Bariadi TC	December	EGR 9-13	Clusters (5)	5	224	WECs, ICs and HTs, st1/2 teachers	33,135,000	147,924	29,585	10.2
G		February	EGR 9-10 review	Clusters (5)	1	86	WECs, ICs and HTs	2,900,000	33,721	33,721	11.6
H		April	EGM 1-4	Residential District	4	125	ICs, HTs, Key Maths teacher, per school and DIC (4 days), HT and WEC (1 day)	25,198,000	201,584	50,396	17.4
I		May	EGM 1-2 review	Clusters (5)	1	115	ICs, Key Maths teacher, st1 teacher, District IC	3,655,000	31,783	31,783	11.0
J	Kilwa DC	December	EGR 9-13 ?	Residential District	4	247	St1/2 teachers?	92,520,000	374,575	93,644	32.3
K		February	Review?	Residential groups (2) & Clusters (10)	1	235	WECs, ICs and HTs	16,540,000	70,383	70,383	24.3
L		April	EGM 1-4	Residential (some 1 day, some 4 days)	4	426	WECs, ICs, HTs, St1/2 teachers, Key Maths teachers (some only 1 day)	94,177,000	221,073	55,268	19.1

Source: LGA spending reports. Notes: Total costs include all allowances, refreshments, fuel, stationary, and District INSET Team facilitator costs. Number of days shows the days of training (not including travel days). *Where some participants only came for part of the training, this is shown in the 'Who were they' column and the Unit cost per day assumed they attended the full session for simplicity (hence the unit cost per day is underestimated). + Costs in GBP calculated using the exchange rate on 30 June 2016 from xe.com.

Annex K   Specialist Evaluation and Quality Assurance Services matrix

Area	Question	Where is this addressed?	Comments
STRUCTURE AND CLARITY: Is the product logically structured, is it clearly written, and does it contain all the relevant elements?	**1.1** Is the product accessible to the intended audience (e.g. free of jargon, written in plain English, logical use of chapters, and appropriate use of tables, graphs and diagrams)?	Volume I and Volume II.	We have published the report in two Volumes in order to achieve the goal of accessibility and methodological rigour. Volume I is the presentation of findings, with minimal technical detail. Volume II is the methodological and in-depth evidence companion. This should enable the intended audiences to access the necessary detail.
	1.2 Is the product of publishable quality?	-	
	1.3 Is it clear who has carried out the evaluation?	Annex B in Volume II	
	1.4 Is an executive summary included, and can it stand alone as an accurate summary of the main product?	Executive summary in Volume I	
	1.5 Do the annexes contain – at the least – the original TORs, the evaluation framework, a bibliography and a list of consultees?	Annex A in Volume I is the Agreed TOR and Annex B in Volume I contains the ML evaluation matrix. References are listed in both Volumes. The acknowledgements in Volume I list individuals consulted.	The ML evaluation matrix is included in the IE ML Planning Report (produced in March 2016), and agreed by DFID and the IE Reference Group.
	1.6 Do annexes increase the usefulness of the product?	-	
	1.7 Have any departures from the original TOR been adequately explained and justified?	Annex A in Volume I describes how and why the TOR evolved.	
CONTEXT: Is the intervention and its policy, development and institutional context clearly described?	**2.1** Does the product provide a relevant and sufficient description of the intervention to be evaluated? At the least, this should include detail on the intervention's anticipated impact, outcomes and outputs, target groups, timescale, geographical coverage, and the extent to which the intervention aimed to address issues of equity, poverty and exclusion.	Annex C.3 in Volume I, the start of each of the Chapters 2,3,4,5, and 6 in Volume I.	

	2.2 Does the product describe the intervention logic and/or theory of change?	Annex C.2 in Volume I and Annex B in Volume I.	Note that developing the expanded TOC was part of the BL work, and is set out in detail in the IE BL report, Volume I Chapter 8.
	2.3 Does the product provide a relevant and sufficient description of the local, national and/or international development context within which the intervention was operating?	Chapter 1 in Volume I and Annex D in Volume I.	The IE ML Planning Report contains further details, and is referenced in the ML IE Report.
	2.4 Does the product identify key linkages between the evaluated intervention and other relevant projects / programmes / donors? If no linkages are identified, does the product justify why other projects / programmes / donors were not relevant to the evaluation?	Chapter 1 in Volume I, Annex D in Volume I, Chapter 3 in Volume II.	The IE ML Planning Report contains further details, and is referenced in the ML IE Report.
	2.5 Is there an assessment of the policy context for the intervention and does this include reference to poverty reduction strategies, gender equality, environmental protection, and human rights?		
	2.6 Does the product describe the extent to which the intervention has been managed and delivered against Paris Declaration principles?	-	This was not within the scope of the evaluation. The TOR does not specify that the evaluation should assess the intervention against the Paris Declaration Principles, and it does not include a process evaluation. Hence the evaluation does not focus on the intervention's management and delivery processes, but it does assess the intervention against the OECD DAC evaluation criteria.
PURPOSE, SCOPE AND OBJECTIVES: Are the purpose, scope and objectives of the evaluation clearly described?	**3.1** Does the product describe what information is needed through the evaluation, and how that information will be used?	Chapter 1 in Volume I and Chapters 2,3,4 and 5 in Volume II explain what information was used and in what way in the evaluation.	
	3.2 Does the product describe whether the evaluation is for accountability and/or learning purposes?	Chapter I in Volume I explains the objectives of the ML stage of the evaluation.	The evaluation objectives show that this evaluation was for both accountability and learning purposes (with emphasis on the former).
	3.3 Does the product describe the target audience(s) for the evaluation?	Chapter 1 in Volume 1 and Annex B in Volume II discusses the desired target audience	

	3.4 Does the product justify the timing of the evaluation?	Chapters 3 and 4 in Volume II explain the rationale for the timing of the research.	The timing of the evaluation is driven by the desire of DFID to have the findings from each round ready in the final quarter of the year when the joint education sector review, and EQUIP-T annual review take place.
	3.5 Does the product clearly outline what aspects of the intervention are and are not covered by the evaluation?	Chapter 1 in Volume I explains the scope of the evaluation.	
	3.6 Are the evaluation's objectives specific and realistic? Are they clearly related to the evaluation purpose?	Chapter I in Volume I explains the objectives of the ML stage of the evaluation.	
EVALUATION METHODOLOGY AND DESIGN: Was an appropriate evaluation framework applied? Were the methods appropriate for addressing the evaluation questions?	**4.1** Is the evaluation framework clearly explained? Does it establish the evaluation questions, data sources and methods for data collection?	Chapter 1 in Volume I and Annex B in Volume I explain the ML evaluation matrix. Data collection methods are set out in Chapters 2, 3, 4 and 5 in Volume II.	
	4.2 Does the product describe and justify which evaluation criteria are applied (e.g. OECD DAC)? Does this include discussion around which criteria were not relevant for this evaluation?	Chapter 8 in Volume I presents findings, against the DAC criteria	The IE ML Planning Report contains further details on the application of the DAC criteria, and is referenced in the ML IE Report.
	4.3 Are the evaluation methods described and justified? Are these methods appropriate for addressing the evaluation questions?	Chapter 1 in Volume I provides a summary. Volume II provides a detailed account of all methodological aspects (Chapters 2,3,4,5 and Annex G).	
	4.4 Is the methodology appropriate for assessing the cross-cutting issues of gender, poverty, human rights, HIV/AIDS, environment, anti-corruption, capacity building, and power relations?		Methods were chosen based on evaluation objectives and questions - gender is a central issue, and poverty is also relevant, in this evaluation and these have been taken into consideration in our sampling, instruments and research methods. Our qualitative research methodology would have allowed us to identify any areas where any other issues would have been of relevance for our research.
	4.5 Is the sampling strategy described, and is it appropriate? Are primary and secondary data sources appropriate, adequate and reliable? Are sample sizes adequate?	Chapter 3 and 4 in Volume II explains the quantitative and qualitative sampling strategies respectively.	The IE BL Report Volume II contains full details of the sampling strategies and is referenced in the ML IE report.

	4.6 Does the design provide for multiple lines of inquiry and/or triangulation of data?	Chapter 2 in Volume II explain the mixed-methods approach and how the different research methods are combined. Each methodological section explains how within-method triangulation was implemented.	
	4.7 Does the methodology enable the collection and analysis of disaggregated data to show difference between groups?		We present results disaggregated by important sub-groups where relevant to our evaluation. Pupil learning results are disaggregated by gender, home language and poverty status.
	4.8 Are any methodological limitations acknowledged and their impact on the evaluation discussed? Are the limitations acceptable and/or are they adequately addressed?	Chapters 3, 4 and 5 in Volume II discusses the limitations for both the qualitative and quantitative aspects of the evaluation and the costing study.	
	4.9 Are any departures from the TOR, inception phase and / or original evaluation design adequately explained?	Chapters 3 and 4 in Volume II explain how and why the ML research differs from the BL research, partly to deal with the evolving programme and education sector context.	
	4.10 Does the product discuss any inherent imbalances or biases that interviews and other data collection may have created?	Chapters 3, 4 and 6 in Volume II.	
	4.11 Does the product describe how any bias has been overcome?	Chapters 3, 4 and 6 in Volume II.	
IMPLEMENTATION: Was the evaluation conducted in an appropriate manner, involving an appropriate range of stakeholders, and taking into account Paris Declaration principles?	**5.1** Were instruments tested and validated (e.g. pre-testing of questionnaires)?	Chapter 3 in Volume II explains the ML pretesting processes.	The IE BL Report Volume II explains the original quantitative instrument design and validation processes, and is reference in the IE ML report.
	5.2 Was data collected in an appropriate and respectful manner, taking into account cultural, ethical and legal concerns?	Annex C in Volume II sets out the ethical principles implemented in the research.	
	5.3 Was there an appropriate level of involvement from the various stakeholders in the design and implementation of the evaluation?	Annex B in Volume II explains the various stages of stakeholder engagement in the ML research.	

	5.4 Did the evaluation process provide affected stakeholders with access to evaluation-related information in forms that respect people and honour confidentiality?	Annex C in Volume II sets out the ethical principles implemented in the research.	
	5.5 Has the evaluation process been transparent enough to ensure its legitimacy?	Annex B in Volume II explains the various stages of stakeholder engagement in the ML research.	
	5.6 Where primary stakeholders were not consulted due to the scope of the evaluation, is relevant documentation drawing on secondary data sources identified and referred to?		Primary stakeholders were consulted via quantitative and qualitative research.
	5.7 Does any summary or description of consultees take into account ethical, privacy and security concerns? (the document should only provide a summary of number and level of staff interviewed, by organisation)	Annex C in Volume II sets out the ethical principles implemented in the research.	
	5.8 To what extent has the evaluation been implemented in accordance with Paris Declaration principles? Have issues of country ownership and management been addressed? To what extent has the evaluation used country systems? How far has the evaluation harmonised approaches with other donors? Has the evaluation contributed to building evaluation capacity within partner countries?	Annex B in Volume II describes how the evaluation has engaged stakeholders; the role of the IE Reference Group which is chaired by the Commissioner for Education and has representatives from key ministries, departments and agencies, the research community, and development partners; and the overall IE governance and management process	The TOR did not specify that the evaluation should be implemented in accordance with the Paris Declaration Principles.
ANALYSIS: Is the analysis sufficiently robust?	**6.1** Is information presented, analysed and interpreted systematically and logically?		
	6.2 Is the analysis presented against the evaluation questions and criteria?	Volume I is structured according to the ML evaluation matrix (which contains research questions linked to the TOC). Chapter 3, 4 and 5 in Volume II explain in detail the sources of data and provide comments on limitations to their quality. In the Findings Chapters in Volume I, important data caveats are explained in the text or footnoted.	
	6.3 Is the evaluation transparent about the sources and quality of information, and are references or sources provided?		

	6.4 Can evidence be traced through the analysis and into findings and recommendations? Is there sufficient cross-referencing? **6.5** Does the analysis include an appropriate reflection of the views of different stakeholders (reflecting diverse interests)? **6.6** Is the analysis disaggregated to show impact and outcomes on the different stakeholder groups? **6.7** Does the analysis explore the cross-cutting issues of gender, poverty, human rights, HIV/AIDS, environment, anti-corruption, capacity building, and power relations?	Chapter 4 in Volume II described the multiple data sources and respondents in the qualitative research, and explains how views of these different types of respondent were considered in the analysis. Chapter 2 in Volume 1 summarises results for different groups of pupils (disaggregation of results by gender, home language and poverty).	Methods were chosen based on evaluation objectives and questions - gender is a central issue, and poverty is also relevant, in this evaluation and these have been taken into consideration in our sampling, instruments and research methods. Our qualitative research methodology would have allowed us to identify any areas where any other issues would have been of relevance for our research.
FINDINGS: Are the findings valid, balanced and adequately supported by evidence?	**7.1** Do findings follow logically from the analysis? **7.2** Do findings address the evaluation questions and criteria? **7.3** Is the relevance of the context (e.g. developmental, policy, institutional) taken into account? **7.4** Is the evidence clear and sufficiently triangulated? **7.5** Are findings useful and are they presented in ways that are accessible to different users?	Chapters 2,3,4,5 and 6 in Volume I are structured according the ML evaluation matrix which contains the evaluation questions.	We do this across the report and particularly present findings with respect to the context of recent education policy reforms. We provide triangulation of findings both within and across methodological components of the evaluation. The division of the report into two Volumes is intended to make the findings accessible to different types of reader.

	7.6 Do findings reflect diverse views and interests? If not, is there adequate explanation for omissions?		The qualitative research uses data from multiple data sources and respondents, and the analysis takes account of different views and interests.
	7.7 Are there appropriate and sufficient findings provided around the cross cutting issues of gender, poverty, human rights, HIV/AIDS, environment, anti-corruption, capacity building, and power relations?		Findings are presented based on evaluation objectives and questions - gender is a central issue, and poverty is also relevant, and these have been taken into consideration.
	7.8 Are issues of attribution considered?	Chapter 1 in Volume I and Chapter 6 in Volume II.	
	7.9 Are unintended and unexpected findings identified?	Examples are found in Volume I Findings Chapters, and in Chapter 8 in Volume I some are referenced in the recommendations.	
RECOMMENDATIONS: Are the recommendations sufficiently targeted and practical?	**8.1** Do recommendations follow logically from the findings and evidence cited?	Chapter 8 in Volume I	Recommendations make reference to our findings.
	8.2 Are they relevant to the evaluation and targeted at the intended users?	Chapter 8 in Volume I	
	8.3 Are they prioritised and clearly presented, enabling individuals or departments to follow up on each specific recommendation?	Chapter 8 in Volume I	
LESSONS: Are the lessons clearly presented and applicable for wider use?	**9.1** Do lessons contribute to general knowledge and are they useful?	Chapter 8 in Volume I	We present lessons together with recommendations in Volume I. At this stage of the evaluation, the focus is on recommendations mainly for the programme to consider for adjustment or consolidation. At the endline, more lessons with wider applicability will be drawn.
	9.2 Are lessons valid (i.e. they have not been generalised from single point findings)?	Chapter 8 in Volume I	
	9.3 Do lessons reflect the interests of different stakeholders, including different sexes?	Chapter 8 in Volume I	
	9.4 Are lessons presented separately with a clear logical distinction between findings, recommendations and lessons learned?	Chapter 8 in Volume I	

USEFULNESS: Has the evaluation process and product adequately addressed the information needs of the commissioning body and other users, and does it address the questions in the TOR?	**10.1** To what extent does the report follow the TOR, and are evaluation questions adequately covered by the report? If not, are departures justified?	Chapters 2,3,4,5 and 6 in Volume I are structured according the ML evaluation matrix which contains the agreed evaluation questions (derived from the TOC).	The ML evaluation matrix is included in the IE ML Planning Report (produced in March 2016), and agreed by DFID and the IE Reference Group.
	10.2 Has the evaluation been designed and managed to meet the information and decision-making needs of the intended users?	Annex B in Volume II explains the various stages of stakeholder engagement in the ML research.	Close communication and interaction with stakeholders and intended users was ensured throughout the implementation of the evaluation. The ML IE Planning Report (March 2016) and the ML IE Preliminary Quantitative Descriptive Findings Report (July 2016) were shared with the IE Reference Group for comment and feedback, which was then considered in the drafting of this ML IE Report. The Draft version of this ML IE report will be presented in workshops with the explicit goal of receiving comments and understanding disagreement.
	10.3 Have stakeholders and end-users been given opportunities to comment on the draft findings, recommendations and lessons? Is there any indication that the evaluation report reflects those comments and acknowledges disagreements?	Annex B in Volume II explains the various stages of stakeholder engagement in the ML research.	
	10.4 Is there any communications plan within the report? If so, does it suggest how dissemination of evaluation results could lead to improved accountability?	Annex B in Volume II explains plans for further stakeholder engagement and dissemination.	
INDEPENDENCE: Is the evaluation sufficiently independent and impartial?	**11.1** Are differences of opinion (within the evaluation team, or amongst stakeholders consulted) fully acknowledged in the report?	Where issues arose, these are discussed. For example, Chapter 3 in Volume II explains how the issue of the risk of contamination was viewed by different parties and the steps taken by the IE team.	
	11.2 Are any conflicts of interest openly discussed?	Annex C.3 discusses potential conflicts of interest, and explains that there were none of relevance to this evaluation. It also discusses the way in which the evaluation team was able to access information freely and work without interference from third parties.	
	11.3 Does the report indicate whether the evaluation team was able to work freely and without interference?		

	11.4 Were information sources and their contributions independent of other parties with an interest in the evaluation?		

About the project

The independent Impact Evaluation of the Education Quality Improvement Programme in Tanzania (EQUIP-T) is a four-year study funded by the United Kingdom Department for International Development (DFID). It is designed to: i) generate evidence on the impact of EQUIP-T on primary pupil learning outcomes, including any differential effects for girls and boys; ii) examine perceptions of effectiveness of different EQUIP-T components; iii) provide evidence on the fiscal affordability of scaling up EQUIP-T post-2018; and iv) communicate evidence generated by the impact evaluation to policy-makers and key education stakeholders.

EQUIP-T is a Government of Tanzania programme, funded by UK DFID, which seeks to improve the quality of primary education, especially for girls, in seven regions of Tanzania. It focuses on strengthening professional capacity and performance of teachers, school leadership and management, systems which support district management of education, and community participation in education.