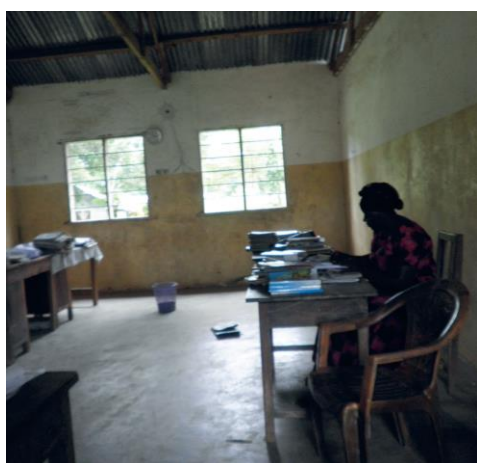# EQUIP-Tanzania Impact Evaluation

Final Baseline Technical Report, Volume II:
Methods and Technical Annexes

Gunilla Pettersson, Georgina Rawle, Rachel Outhred, Stephanie Brockerhoff, Gabrielle Wills, Dita Nugroho, Paul Jasper, Andrej Kveder and Adrian Beavis

January 2015

Acknowledgements

Table of contents

List of figures and tables

List of abbreviations

BRN	Big Results Now
CAPI	Computer-Assisted Personal Interviewing
CI	Confidence interval
DEO	District Education Officer
DFID	Department for International Development
DSI	District School Inspector
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EMIS	Education Management Information System
EQUIP-T	Education Quality Improvement Programme in Tanzania
FGD	Focus Group Discussion
HH	Household
ICT	Information and Communication Technology
IE	Impact Evaluation
IEC	Information, Education and Communication
ICC	Intra Cluster Correlation
IGA	Income-Generating Activity
INSET	In-service Teacher Training
KII	Key Informant Interview
LANES	Literacy and Numeracy Education Support Programme
M&E	Monitoring and Evaluation
MA	Managing Agent
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
PMO-RALG	Prime Minister's Office Regional Administration and Local Government

PSA	Programme Support Activities
PSLE	Primary School Leaving Examination
PSM	Propensity Score Matching
PSM-DID	Propensity Score Matching-Differences in Differences
PTG	Parents–Teachers Group
PTR	Pupil–Teacher Ratio
RCT	Randomised Control Trial
REO	Regional Education Officer
RTI	Research Triangle Institute International
SACMEQ	Southern African Consortium for the Measurement of Education Quality
SC	School Committee
SDIS	Service Delivery Indicators Survey
SES	Standardised Effect Size
SEQAS	Specialist Evaluation and Quality Assurance Services
SLM	School Leadership and Management
SP	Service Providers
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
TPMS	Teacher Performance Management System
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNICEF	United Nations Children's Fund
WEC	Ward Education Coordinator
WSDP	Whole School Development Plan

PART E: INTRODUCTION

1.1   Overview

The Education Quality Improvement Programme in Tanzania (EQUIP-T) is a four-year, Government of Tanzania programme, funded by the UK DFID that will seek to improve the quality of education, especially for girls, in seven regions in Tanzania (Cambridge Education 2014). The independent impact evaluation (IE) conducted by Oxford Policy Management (OPM) will measure the impact of the EQUIP-T programme as a whole on selected education outcomes that the programme will seek to influence over time. For details on the scope of the IE, see the terms of reference (TOR) in Volume I, Annex A.

This report is organised into two volumes. Volume I presents an overview of the evaluation, the baseline results for programme treatment districts, and an assessment of the EQUIP-T programme TOC. It is designed to be accessible to all readers. Volume II (this document) covers the technical and methodological detail, and is intended for those interested in methods, detailed statistical results, and the detailed qualitative background paper.

This report will also be supplemented by stand-alone policy briefs on teacher capacity and performance and pupil learning.

1.2   Structure of this Volume

This is Volume II of the EQUIP-T Impact Evaluation Baseline Report, which accompanies *Volume I: Results and Discussion*. It is structured as follows.

Part B begins by setting out the mixed methods approach in chapter 2. Next, chapter 3 describes the quantitative methodology including identification of the control group; the sampling strategy, sample size and survey weights; the survey instruments, impact indicators and impact measurement for the subsequent rounds. The chapter finishes with a discussion of risks to and possible limitations of the quantitative component of the IE. It highlights one particular emerging risk of contamination from another large education programme, and the steps being taken to mitigate this. The qualitative research methodology is discussed in chapter 4 starting with the research objectives before the selection of qualitative research sites, instruments and tools, fieldwork and analysis to be used are discussed. The chapter ends with an overview of possible limitations of the qualitative research design.

The supplementary results and technical annexes of this volume contain the following: map of the IE districts (Annex A), overview of stakeholder baseline consultations (Annex B), description of permits and consent seeking for the fieldwork, confidentiality and datasets (Annex C), dissemination (0), quantitative data collection, cleaning and analysis (Annex E), definitions of key quantitative indicators (Annex F), supplementary quantitative baseline results (Annex G), detailed tables for the Part B results (Annex H), details on the pupil tests including Rasch analysis (Annex I), details on the qualitative sampling, tools and fieldwork (Annex J), qualitative evaluation matrix (Annex G), key qualitative research questions and data sources (Annex L), qualitative analysis background paper (Annex M), a summary of the concept note for the separate fiscal study that will accompany the IE (Annex N), a description of the IE governance structure (Annex O), and the SEQAS matrix (Annex P).

PART F: Methods

2   Mixed methods approach

The EQUIP-T Impact Evaluation Inception Report outlines the strong case for a mixed methods approach to impact evaluation (OPM 2014a). The three main considerations in designing the mixed methods approach included: integration of methodologies for better measurement; sequencing information for better analysis; and merging findings for better action (Carvalho and White 1997; Garbarino and Holland 2009). Eight steps were proposed to integrate methodologies for better measurement; six suggested steps to sequence information for better analysis; and one final step to merge findings at analysis stage into a single, mixed methods report (Garbarino and Holland 2009). The IE uses 14 of the 15 steps proposed by Garbarino and Holland (2009).

Specifically, the IE does the following:

Integrates methodologies for better measurement

- Uses indicators from existing education surveys and quantitative data sources and the baseline quantitative survey to select a qualitative investigation sample;

- Uses quantitative data to highlight priority issues and generate new hypotheses to test through qualitative research;

- Uses qualitative analysis to identify knowledge gaps to be filled by the quantitative survey;

- Uses qualitative analysis to prioritise issues important to stakeholders that should be covered by a survey;

- Uses qualitative analysis to construct indicators;

- Uses qualitative analysis to define sub-group sampling frames; but

- Does not use qualitative analysis to stratify the quantitative sample.

Sequences information for better analysis

- Generates working hypotheses from an initial qualitative study to test using quantitative methods;

- Uses a qualitative study in a sub-sample of quantitative areas to elicit rich contextual data that may affect programme outcomes;

- Uses a qualitative study to explore contextual variation that may result in heterogeneous programme outcomes;

- Uses the qualitative study to explain relationships emerging from the quantitative survey;

- Uses the qualitative study to compare survey findings where possible; and

- Uses the qualitative study to enrich analysis of relationships/trends/patterns from the quantitative survey.

Merges findings at analysis stage into a single, mixed methods report

- Produces an integrated mixed methods report drawing on both quantitative and qualitative data.

2.1   Combining qualitative and quantitative data

Figure 1 illustrates how the quantitative and qualitative data sources will be combined to yield an understanding of (any) EQUIP-T impact and potential mechanisms through which the programme may have influenced the outputs and outcomes of interest. It shows that while the quantitative survey focuses on collecting data at the school level covering head teachers, teachers and pupils in a representative sample of programme treatment schools, the qualitative research collects data in a small purposive sample of districts, wards, communities and schools and from EQUIP-T programme staff at regional level. The quantitative and qualitative components are designed to complement each other to allow for deeper insights into any impact and mechanisms of EQUIP-T[1].

Figure 1 Quantitative and qualitative data collection
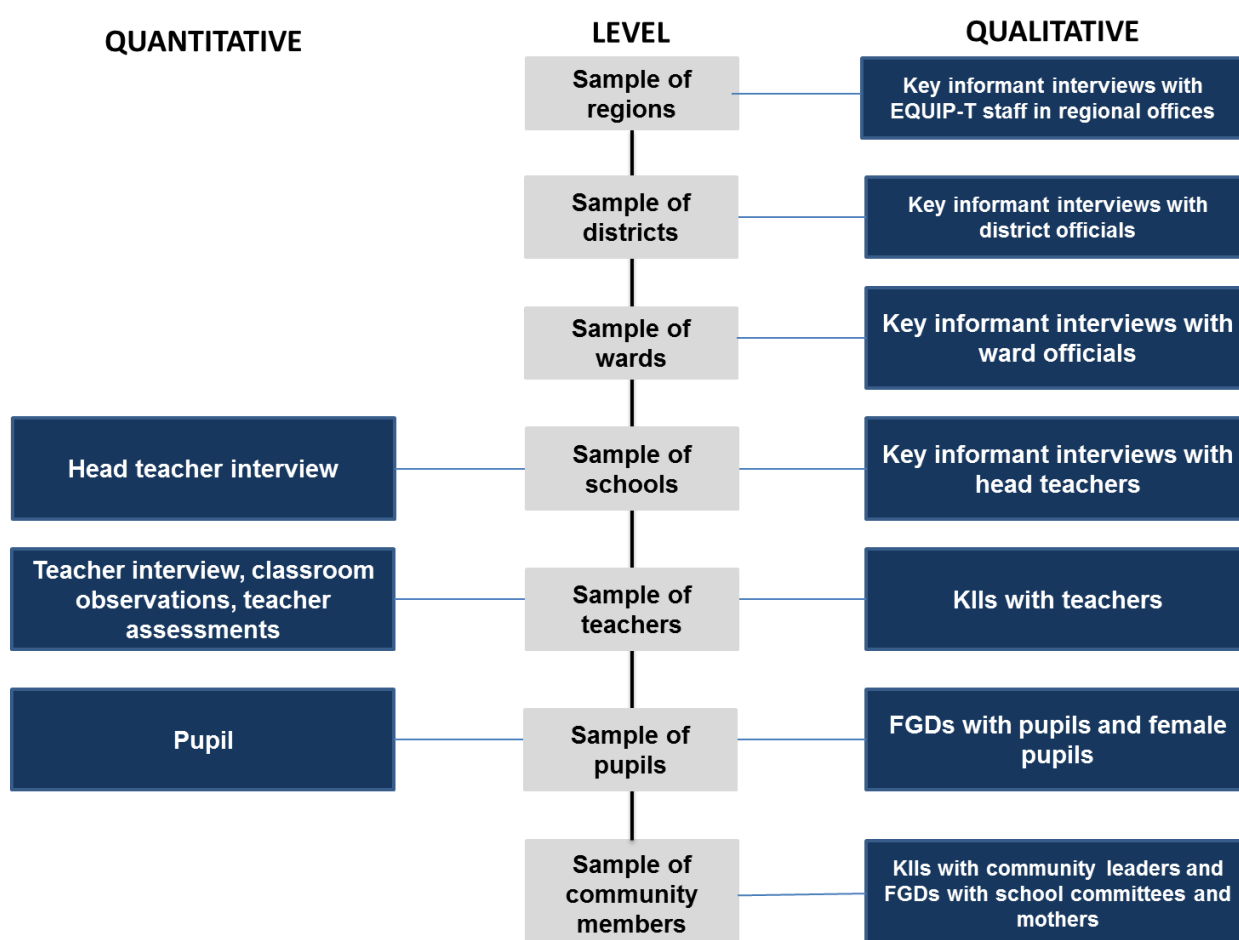
Source: OPM impact evaluation team.

The timing of the qualitative fieldwork was adjusted so that the first part would take place in early July 2014 (regional, district and ward level KIIs) and the second part in July-August 2014 (school and community KIIs and FGDs) after the completion of the quantitative fieldwork. This was to allow the qualitative fieldwork to explore selected initial findings from the quantitative research with head teachers, teachers and pupils. As such, part of the qualitative research programme was left open to be flexible to the emerging findings from the quantitative survey.

---

[1] The IE also contains a fiscal study with the overall objective of estimating the cost effectiveness of EQUIP-T and assess the affordability of scaling up EQUIP to remaining (rural) districts in Tanzania after 2018. This is a separate study. For a summary of the fiscal study concept note see Annex N.

The quantitative data collection and initial analysis took place before the qualitative design was finalised, in order that the findings from the quantitative baseline could feed into the qualitative research design to enable it to explore issues uncovered by the preliminary quantitative data analysis and factors that contribute to observed outcomes at school level. Specifically, the qualitative research guides for the school and community qualitative research were finalised after the completion of the quantitative baseline survey, with some questions based on the preliminary quantitative findings.

3 Quantitative methods: Quasi-experimental design

This chapter starts by describing the methodology for the quantitative approach including assessment the impact of the programme as a whole, identification of the control group, the sampling strategy and sample size. It then discusses the development of the survey instruments and discusses each instrument before describing the impact indicators and the approach to impact estimation in 2016 (2018), and the follow-up survey. Finally, it outlines risks to and possible limitations of the quantitative approach.

3.1 Methodology

The key challenge in evaluating the impact of any programme is the identification of an appropriate control group that does not benefit from the programme, and which can be used as a counterfactual to the treatment group that does benefit. Provided one has been successful in selecting an appropriate control group, the post-programme differences in the outcomes of the treatment and control groups can be used to estimate the causal impact of the programme that is being evaluated.

In an ideal situation for the EQUIP-T IE, the selected control group would be a 'perfect clone' of the treatment group, with both groups being statistically identical with the only difference between the two groups being exposure to the programme. In reality this is not possible and the main problem faced in identifying a control group that can act as the appropriate counterfactual is selection bias.

Selection bias occurs if there is some set of characteristics that are non-randomly different across the treatment and control groups, such that: (1) there is an increase in the likelihood of a particular type of individual (e.g., pupil, teacher or head teacher) being exposed to the programme; and (2) these characteristics influence outcomes against the key impact indicators. If this is the case then the IE would be unable to distinguish the true impact of exposure to the EQUIP-T programme, from the non-random characteristics that drive the probability of exposure of individuals to the programme in the first place.

To understand how selection bias could potentially affect the IE, imagine a situation where treatment and control groups are selected such that even in a pre-programme district, pupils in the control group are attending schools that are performing worse as compared to the schools attended by pupils in the treatment group. In such a scenario the IE would be unable to determine whether any differences in key outcome indicators across treatment and control pupils are due to the programme or pre-existing differences in performance across treatment and control schools.

In order to provide a framework for evaluating the appropriateness of any proposed methodology it is useful to consider the three conditions that an appropriate counterfactual must satisfy (Gertler et al. 2011):

- Treatment and control groups should share on average the same characteristics;

- Treatment and control groups should react to the programme in the same way; and
- Treatment and control groups should not be differentially exposed to other interventions during the period of the evaluation.

3.1.1 Assessing the impact of the EQUIP-T programme as a whole

The EQUIP-T programme regions and districts were purposively selected by the MA on the basis of region and district rankings in terms of education performance and financial resources. The selection criteria for regions were: Primary school leaving examination (PSLE) pass rates, primary pupil-teacher ratios (PTR) and female lower secondary enrolment, and for districts: primary education spending per capita, Kiswahili literacy rates for standard 3 pupils, primary PTRs and female lower secondary enrolment for districts. Regions and districts with relatively weaker education performance and resources were selected by the EQUIP-T programme. Within the EQUIP-T districts, all government schools will be covered by the programme.

Thus a pure randomised control trial (RCT) approach was not an option for the IE. A staggered EQUIP-T programme roll-out across treatment districts was not possible either as DFID would like all programme districts to receive the same combination of EQUIP-T interventions without any timing differential. In the absence of random assignment, the IE employed the best possible approach to simulate the RCT approach – in this case this was to mimic randomisation by use of propensity score matching (PSM). To measure EQUIP-T programme impact as a whole on key outcomes and outputs (see sections 3.4 and 3.5), the IE will use a PSM-DID approach.

3.1.2 Identifying the control group

In terms of assessing the impact of the EQUIP-T programme as a whole, PSM was used to construct a counterfactual to address the problem of selection bias. A two-stage sampling process was used. A counterfactual (control group) was constructed by first matching control districts to the pre-selected treatment districts on observable characteristics using propensity score matching (PSM). Second, schools were randomly sampled within the treatment districts and control schools were matched to these using PSM. For the matching, existing secondary data sources including Education Management Information System (EMIS) data and UWEZO learning assessment micro data were used.

Thus, the IE approach will allow for the robust estimate of the impact of the EQUIP-T programme as a whole over time by comparing differences in outcomes between the treatment and matched control groups (taking account of starting points as required), using a PSM-DID approach (for risks and limitations see sections 3.5.2 and 3.7). But the IE will not quantitatively measure the relative effectiveness of different EQUIP-T components given the quantitative IE design option chosen by DFID (see Volume I, Annex A.2.1.2 and A.2.1.3).

The first EQUIP-T activities will be rolled out in July 2014 and the IE midline survey will take place in March-May 2016. Two years may be a relatively short time after which to expect any impact on pupil learning, but the IE will also measure changes in EQUIP-T programme outputs such as teacher capacity and performance, which would be expected to change more quickly than pupil learning outcomes, if the EQUIP-T programme works as intended.

3.2   Sampling strategy and sample size

Because EQUIP-T regions and districts were purposively selected (section 3.1.1), the IE sampling strategy used PSM to: (i) match eligible control districts to the pre-selected and eligible EQUIP-T

districts (for eligibility see Box 1), and (ii) match schools from the control districts to a sample of randomly sampled treatment schools in the treatment districts.

The same schools will be surveyed for each round of the impact evaluation (panel of schools) but there will not be a panel of pupils (also see below).

Box 1 Identifying districts eligible for matching

Eligible control and treatment districts were those not participating in any other education programme or project that may confound the measurement of EQUIP-T impact. To generate the list of eligible control and treatment districts, all districts that are contaminated because of other education programmes or projects or may be affected by programme spill-over were excluded as follows:

- *All districts located in Lindi and Mara regions as these are part of the EQUIP-T programme, but the IE does not cover these two regions[2];*

- *Districts that will receive partial EQUIP-T programme treatment[3] or will be subject to potential EQUIP-T programme spill-overs;*

- *Districts that are receiving other education programmes and projects that aim to influence the same outcomes as the EQUIP-T programme and which could confound measurement of EQUIP-T impact: BRN, Kiufunza, UNICEF programme and USAID TZ21 programme;*

- *Districts that were part of OPM's IE pre-test 1 (two districts); and*

- *Districts that were part of OPM's IE pre-test 2 (one district).*

3.2.1 Sampling frame[4]

To be able to select an appropriate sample of pupils and teachers within schools and districts, the sampling frame[5] consisted of information at three levels:

- District level
- School level
- Within school level

The sampling frame data at the district and school levels was compiled from the following sources: the 2002 and 2012 Population Censuses, EMIS data from MoEVT and PMO-RALG, and UWEZO 2011 learning assessment micro data.

For within school level sampling, the frames were constructed upon arrival at the selected schools and was used to sample pupils and teachers on the day of the school visit.

3.2.2 Sampling stages

To select the treatment and control schools in a way that meet the counterfactual conditions discussed in section 3.1, a four-stage approach was used.

---

[2] The programme will start one year later in districts in Lindi and Mara regions than in the other programme districts.

[3] Most likely all districts in EQUIP-T programme regions will receive some type of intervention according to discussions with the MA in November, 2013.

[4] A separate note on the district and school sampling frame is available from OPM on request.

[5] One of the challenges during the sampling frame construction were changes in the administrative boundaries in Tanzania; the biggest and most recent change occurred before the 2012 Population Census. To enable merging of data from different years (2011, 2012 and 2013) a correspondence table was developed that can be used to map district level indicators from different years despite their different administrative divisions.

Stage 1: Selection of control districts

Because the treatment districts were known, the first step was to find sufficiently similar control districts that could serve as counterfactuals.

PSM was used to match eligible control districts to the pre-selected, eligible treatment districts using the following matching variables: Population density, proportion of male headed households, household size, number of children per household, proportion of households that speak an ethnic language at home, and district level averages for household assets, infrastructure, education spending, parental education, school remoteness, pupil learning levels and pupil drop out.

Stage 2: Selection of treatment schools

In the second stage, schools in the treatment districts were selected using stratified random sampling. The schools were selected using a probability proportional to size approach, where the measure of school size was the standard two enrolment of pupils. This means that schools with more pupils had a higher probability of being selected into the sample. To obtain a representative sample of treatment schools, the sample was implicitly stratified along four dimensions:

- Districts;
- PSLE scores for Kiswahili;
- PSLE scores for mathematics; and
- Total number of teachers per school.

Stage 3: Selection of control schools

As in stage one, a non-probabilistic PSM approach was used to match eligible control schools to the sample of treatment schools. The matching variables were similar to the ones used as stratification criteria: Standard two enrolment, PSLE scores for Kiswahili and mathematics, and the total number of teachers per school. This way, not only were the treatment districts and control districts as similar as possible (given the available data), but also the schools within these districts.

The midline and endline surveys will be conducted for the same schools as the baseline survey (a panel of schools). However, the IE will not have a panel of pupils as a pupil only attends standard three once (unless repeating). Thus, the IE will have a repeated cross-section of pupils in a panel of schools.

Stage 4: Selection of pupils and teachers within schools

In stage four, pupils and teachers were sampled within schools using systematic random sampling (SRS) using school registers. Per school, 15 pupils were sampled[6]. In addition teachers were sampled at each school: Up to three teachers who teach standards 1-3 mathematics, up to three teachers who teach standards 1-3 Kiswahili, and up to three teachers who teach mathematics to standards 4-7[7]. Often, teachers in standards 1-3 teach both Kiswahili and mathematics. If this was the case, these teachers were sampled for both the Kiswahili and mathematics teacher development needs assessment (TDNA). The within school sampling was assisted by selection tables automatically generated within the computer assisted survey instruments.

---

[6] This number was obtained through a theoretical power calculation (see OPM 2014a).

[7] These teachers were not necessarily those who teach the sampled standard three pupils. Although in many schools this would be the case as the majority of schools are relatively small.

Replacement sample

If a selected school could not be surveyed it was replaced. In the process of sampling, the OPM IE evaluation team drew a replacement sample of schools, which were used for this purpose (reserve list) and the use of this list was carefully controlled. Five out of the 200 original sample schools were replaced during the fieldwork (see Annex E).

To preserve the validity of the IE results the list of districts and schools selected for the baseline survey will be kept confidential. The OPM IE team will therefore not share this list until 2018 after the completion of the endline survey.

3.2.3 Sample size

The required sample size depends mainly on the expected changes to be brought about by the EQUIP-T programme, which according to the EQUIP-T Intervention Summary was to raise the proportion of standard three pupils proficient in Kiswahili from 23% to 50% at the end of the programme (DFID 2013)[8]. However based on existing evidence and OPM's experience in evaluating large-scale, multi-year education quality improvement programmes in other countries, achieving more than a doubling in the proportion of children proficient in a subject in in the span of two to four years is ambitious.

The sample size required to detect change depends on the impact; the smaller the change, the larger the number of pupils required. The larger the number of pupils tested, the easier it will be to detect any changes. This can be achieved by increasing the number of schools, by increasing the number of pupils within each school, or both.

Pupils in the same school tend to be more similar to each other than to pupils from other schools. Thus, an increase the number of schools rather than an equivalent increase in the number of pupils within each school would increase the ability to detect change by a larger factor[9]. That is, an increase in the number of sample schools would increase the IE's ability to detect smaller changes in pupils' knowledge proficiency more than an equivalent increase in pupils within schools. Based on the above considerations, OPM's technical proposal proposed a minimum sample of 100 treatment and 100 control schools, but strongly recommended a sample size of 150 treatment and 150 schools, but the smaller sample size was selected by DFID (see Volume I, Annex A.2.1.3).

From the 200 selected schools, the IE sampled 15 standard three pupils per school to be tested in both Kiswahili and mathematics, balanced by gender (Table 1). Standard three pupil learning levels in Kiswahili and mathematics respectively are the two key indicators of the IE (see section 3.4). As the sample will by design contain relatively small schools, the OPM IE team examined early grade enrolment data for the EQUIP-T districts and the vast majority of schools have more than 20 pupils in standard 2 (EMIS 2013), which meant that 15 standard three pupils could be sampled per school.

The IE also sampled and interviewed up to three teachers who teach standards one to three Kiswahili and/or mathematics, and assessed (see section 3.3.2) up to three standard one to three teachers who teach Kiswahili and/or mathematics, and up to three standard four to seven teachers who teach mathematics per school[10].

---

[8] DFID (2013). The EQUIP-T programme also expects to improve pupil mathematics outcomes.

[9] To control for this effect the OPM IE team assumed a degree of similarity between the children's education outcomes who attend the same schools, Intra Class Correlation (ICC), to be approximately 15% for all the sampling considerations.

[10] Teachers who teach standards four to seven were not interviewed.

It is important to note that the three survey instrument pre-tests (see section 3.3) conducted in schools similar to those in the treatment districts showed that it in many cases there are only three or fewer teachers who teach standards one to three per school, and similarly for teachers who teach standards four to seven. This limitation in terms of the teacher sample size was beyond the IE design given the school sample size chosen by DFID and was set out in the inception report (OPM 2014a).

Table 1 shows the intended and actual sample sizes for the quantitative baseline survey. In general the unit response rates were high: 200 schools in the 25 districts were surveyed and 2,987 standard 3 pupils were tested (99.6% of intended sample size) and poverty scorecards for tested pupils' parents were administered for 2,893 of the pupils (96.4% of intended sample).

As discussed above, many schools in reality had three or fewer teachers who teach standards one to three, and therefore the theoretical power calculations (OPM 2014a) were conducted assuming two and three teachers respectively, for minimum intended sample sizes of 400 (small sample size) and 600 (large sample size) respectively. The actual number of interviews with teachers of standards one to three was 681, more than the intended 600. For the teacher development needs assessments (TDNA) in Kiswahili and mathematics, 510 teachers who teach standards one to three Kiswahili (85% of intended large sample size); 505 teachers who teach standards on to three mathematics (84% of intended sample large sample size); and 564 teachers who teach standards four to seven mathematics (94% of intended large sample size) completed the TDNA for the appropriate subject. These non-responses were due to there being fewer than three teachers at the school (either because it was a very small school or because of teacher absence). Only in a few isolated cases did a teacher refuse to complete the assessment. For the TDNA mathematics there is also the option of combining teachers who teach mathematics to standards one to three with those who teach it to standards four to seven for a combined sample size of 1,023[11].

Table 1 Overview of intended and actual sample sizes

Sampling unit	Treatment intended sample size per unit	Control intended sample size per unit	Total intended sample size	Treatment actual sample size per unit	Control actual sample size per unit	Total actual sample size	Unit response rate (%)
Districts total	17[1]	8[2]	25	17	8	25	100.0
Schools total	100	100	200	100	100	200	100.0
St. 3 pupils per school tested in Kiswahili and mathematics	15	15	3000	15.0	14.9	2987	99.6
Poverty scorecard for tested pupils' parents	15	15	3000	14.4	14.5	2893	96.4
Teachers of stds 1-3 Kiswahili and mathematics interviewed	up to 3	up to 3	up to 600	3.3	3.5	681	113.5
Teachers of stds 1-3 administered Kiswahili TDNA	up to 3	up to 3	up to 600	2.5	2.6	510	85.0
Teachers of stds 1-3 administered mathematics TDNA	up to 3	up to 3	up to 600	2.5	2.6	505	84.2
Teachers of stds 4-7 administered mathematics TDNA	up to 3	up to 3	up to 600	2.8	2.8	564	94.0
St. 2 lesson observations in Kiswahili and mathematics	2	2	400	1.99	1.98	397	99.3

Note: (1) Eligible treatment districts (excluding districts in Lindi and Mara regions as these two regions are not covered by the IE); (2) Eligible control districts; (3) Lesson observations were not sampled.

---

[11] This sample size is 1,023 because some teachers teach both standards one to three and four to seven.

Item response rates

Item response rates were generally very high. Exceptions include: Actual school open days (11% missing), pupil age, which was self-reported by pupils (16% missing); and capitation grant payments per pupil expected in 2012 and in 2013 respectively (32% missing). For the intended and actual number of observations for each of the indicators presented in Volume I, Part B, see Annex H and for TDNA results adjusted for non-response see Annex table 24 and Annex table 25.

3.2.4   Survey weights

In order to obtain results that are representative of the EQUIP-T programme areas that are part of this IE, estimates for treatment areas in this report were weighted using survey weights that were normalised values of the inverse probabilities of selection into the sample for each unit of observation. The relevant probabilities of selection differ depending on whether analysis is carried out at school, pupil, or teacher level and survey weights were calculated for each of these levels.

School weights

Treatment was purposively assigned at the district level and was not influenced by the sampling process. Hence, sampling started at the within-district level, where schools in treatment districts were selected depending on their size in terms of pupils enrolled from among all of the schools in the eligible EQUIP-T programme areas. Therefore, the probability of being selected of each school depended on the total number of schools being selected and, crucially, its size relative to the total number of enrolled pupils across all schools in the treatment areas. Formally, this can be expressed as follows:

$$p_i^{\text{s}} = \frac{mn_i}{N};$$

where p_i^{s} is the probability of school i to be selected into the sample, m is the total number of schools sampled (100 in the present case), n_i is the number of pupils in school i given by the sampling frame data, and N is the total number of pupils in all schools in the relevant programme areas. School-level weights were appropriately normalised inverses of these probabilities.

Pupil weights

At the pupil level, the probability of selection of each pupil was given as follows:

$$p_i^{\text{p}} = p_i^{\text{s}} \cdot \frac{b}{B_i};$$

where p_i^{p} is the probability of each pupil in school i to be selected, p_i^{s} is defined as above, b is the number of pupils selected per school (15 in the present case), and B_i is the total number eligible pupils in school i identified during field work. Again, weights are normalised inverses of these probabilities.

Teacher weights

In principle, teacher level weights should vary depending on the 'type' of teacher being considered. For example, for TDNA Kiswahili indicators, the probability of being selected into the sample would depend on the total number of Kiswahili teachers in standards 1-3 per school. Similarly, for Mathematics TDNA indicators, the probability of being selected into the sample would depend on the number of mathematics teachers in the appropriate standards per school.

However, calculating different teacher level weights for each of the 'types' of teacher would make defining the analysis set-up very complex and prone to errors. In addition, since the total number of eligible teachers relative to the numbers sampled was relatively high in most schools, weights would not vary much across the different teacher types. Therefore, one type of teacher level weight was calculated for the purposes of this report, which takes into account the size of each school in terms of teachers that could be selected into either of the two TDNAs. Hence, the probability of selection for teachers can be expressed as follows:

$$p_i^t = p_i^s \cdot \frac{t_i}{T_i};$$

where again p_i^s is the school probability of selection and p_i^t is the probability of selection for teachers in school i. The denominator T_i is the total number of teachers that were identified to be eligible for either of the TDNA tests in school i, i.e. the total number of teachers that taught mathematics to standards 1-3, mathematics standards 4-7, and/or Kiswahili to standards 1-3 respectively. The numerator t_i is the number of teachers that were actually selected for any of the instruments per school, i.e. if they were administered any TDNA. Note that in contrast to the number of pupils selected per school, this value was not the same across all schools, but depended on the composition and size of the teaching body in each school. As before, weights were the normalised inverse of p_i^t. These weights were applied to all instances where teacher level indicators from the TDNAs and interviews were estimated.

The TDNA data is based on a sample of teachers. When teacher level indicators (for instance, teacher absenteeism) were instead constructed based on the roster (a census of all teachers in the sampled schools), school level weights were used. This is because all teachers in each school are included in the roster, which means that the probability of selection for each teacher is equal to one for these indicators.

Estimation set-up

These weights were used within a survey set-up in Stata that took into account clustered sampling, stratification, and finite population corrections.

Clustering was only relevant for pupil and teacher level indicators, as schools were the primary sampling units within the eligible treatment districts. School pupil data is also hierarchical in nature with pupils clustered within schools. Hence, for pupil and teacher estimates, clustering was set at the school level.

Stratification during sampling was used at the primary sampling level, that is, at school level, and not at lower levels (pupil and teacher). For the estimation set-up, strata for schools were defined by districts and teacher-body size terciles. Although, during sampling, schools were implicitly stratified by PSLE examination scores as well, this is a continuous variable that cannot be used to define strata in the estimation set-up.

Finally, as large proportions of the total eligible population were sampled in many schools at the teacher and pupil levels, the estimation set-up also accounted for the finite population correction (FPC) factor. This factor is generally defined as follows:

$$FPC = \sqrt{\frac{N-n}{N-1}};$$

where N is the size of the population from which the sample is drawn and n is the size of the sample. Hence, in the case of school level indicators, this FPC factor is constant across all

schools, as the sample of schools was drawn from a constant population of all eligible schools in treatment areas. For teacher and pupil level indicators, however, this FPC factor changes depending on the school, as population sizes and, in the case of teacher level indicators, sample sizes vary as well.

In order to understand how large a change the IE design will be able to detect theoretical power calculations were provided in the EQUIP-T Impact Evaluation Inception and Baseline Design Report (OPM 2014a). Based on these calculations, assuming that the proportion of standard 3 pupils proficient in Kiswahili was measured at 50% at the onset of EQUIP-T, the sample size would enable detection of a minimum eight percentage point increase in proficiency from 50% to 58% of pupils (equivalent to a 16% increase in the proportion of pupils proficient in Kiswahili). However, if the initial proportion of pupils proficient in Kiswahili was instead 15%, then the sample size would allow detection of a minimum six percentage point increase from 15% to 21% (40% increase in the proportion of pupils proficient in Kiswahili).

Revised power calculations based on the baseline and follow-up survey data will be contained in the 2016 EQUIP-T Impact Evaluation Midline Report. The actual power of detection for the pupil and teacher level indicators respectively will not be known until the revised power calculations are done in 2016. It should be noted that the school sample size selected was the minimum one required (see Volume I, Annex A.2.1.2 and A.2.1.3), which means the power for teacher level indicators may be low as there were generally relatively few teachers per school as the schools were in general small and remote (also see section 3.2.3).

3.3   Survey instruments

The IE administered eight different instruments in the sampled primary schools using Computer-Assisted Personal Interviewing (CAPI) except for the teacher development needs assessment (TDNA) instruments, which were administered on paper as these take the form of mock pupil tests in Kiswahili and mathematics (Table 2).

Table 2 Quantitative survey instruments

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)
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. 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
4. 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)
5a. 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)
5b. 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)
6. Head teacher interview and data collection from school records		
•	Background information on head teacher: gender, age, years of experience, qualifications	Head teacher interview and school records checks
•	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 and by headcount on the day)	
•	Pupil attendance (by records and by headcount on the day)	
7. Standard 2 Kiswahili and mathematics lesson observations		
•	Inclusive behaviour of teachers with respect to gender	Lesson observation
•	Key teacher behaviours in the classroom	
•	Availability of lesson plan	
•	Availability of seating, textbooks, exercise books, pens/pencils etc. during the lesson	

3.3.1 Development of survey instruments

To develop the quantitative survey instruments the OPM IE team collected and reviewed relevant existing survey instruments, discussed with national and international experts the different types of instruments, visited schools during the November 2013 and January 2014 missions (see Annex table 2), worked with a national team of experts to develop the pupil assessments and the teacher development needs assessments (TDNA), and conducted three pre-tests of the instruments.

The first pre-test (January 2014) tested the initial set of draft instruments before starting to programme these into CAPI, and took place in eight primary schools in two districts (Kisarawe and Mkuranga) in Pwani region. This was conducted by a national team consisting of Kiswahili and mathematics specialists from the University of Dar es Salaam, primary school teachers and an expert who had previously developed learning instruments for the UWEZO learning assessment and Kiufunza project in Tanzania, accompanied by two OPM IE team members. The instrument revisions made based on the pre-test 1 findings are discussed in the relevant sections below. The aims of this pre-test were to:

- Assess whether the difficulty range of the items in the draft pupil tests corresponds to the Tanzania primary 1 and 2 curriculum levels for each subject (Kiswahili and mathematics);
- Test whether the draft pupil test items discriminate adequately among pupils (therefore poorly and strongly performing schools were purposively selected on the basis of PSLE scores);
- Assess whether the difficulty range of the draft items in the draft TDNAs corresponds to the Tanzania primary school curriculum for each subject (Kiswahili and mathematics);
- Test whether the draft TDNA items discriminate adequately among teachers;
- Test appropriate time of administration for the pupil tests and TDNAs; and
- Test accuracy of all instruments in terms of precision and clarity of wording, translation, answer categories etc.

The second pre-test (February 2014) was conducted in four schools in Morogoro DC in Morogoro by two teams consisting of local enumerators and OPM IE team members. The aims of this pre-test were to:

- Test accuracy of the instruments in terms of precision and clarity of wording, translation, answer categories etc.;
- Test technical aspects of administering the instruments using CAPI;
- Test instruments and procedure from an education perspective;
- Test and improve fieldwork protocols and logistics;
- Learn about the administration time of the instruments for fieldwork planning purposes:
- Test technical design issues (reliability of pupils' assessment on household socio-economic status and consistency with data analysis plan);
- Assess adequacy of materials and equipment;
- Learn about potential challenges in the field;
- Gather field experience to improve training delivery;
- Further test appropriate time of administration for the pupil tests; and
- Test whether it is appropriate to administer both tests to the same pupil.[12]

The third and final pre-test was conducted in March 3-5, 2014 by a team of OPM IE team members and local enumerators to test the revised CAPI software, gain further insight for the enumerator training and collect additional information to inform the baseline fieldwork planning. The instruments were then further revised, in particular, translations, during the enumerator training in mid- to late March, 2014.

3.3.2 Instrument descriptions

The overall key intended impact of EQUIP-T is improved pupil learning. Accordingly, the OPM IE team has attached particular importance to the measurement of pupil learning and the development of the instruments intended to do this is described in some detail below. In regard to the development of each instrument, while the comprehensive pre-testing process has been described earlier, it should also be noted that the Reference Group for the Impact Evaluation was provided with penultimate drafts of survey instruments and its comments guided the instrument revisions and finalisation.

Pupil learning assessment

The IE administered EGRA and EGMA style assessments developed in Tanzania specifically for the IE to standard 3 pupils. These instruments capture core standard 1 and 2 curriculum skills in Kiswahili literacy: reading, reading comprehension and listening comprehension and writing, and in mathematics.

The learning assessment instruments developed for the IE will be used to test standard 3 pupils in 2016 and 2018 as well. The reason for testing standard 3 pupils instead of standard 2 and standard 5 pupils, is that the EQUIP-T programme will target the early grades in the first phase, with the aim of improving standard 3 pupil learning (EQUIP-T logframe outcome indicator 1 see Cambridge Education 2014).

In the technical proposal the UWEZO learning assessment instrument was proposed to be used to measure pupil learning and this possibility was discussed with UWEZO Tanzania and UWEZO's head office in Nairobi, who were extremely helpful and willing to share information. However, the UWEZO learning assessment is designed to respond to different information needs from those of the IE (to measure EQUIP-T impact). Further, the UWEZO learning assessment instruments are

---

[12] The suggestion to test the same pupil in both subjects was made by the Reference Group for the Impact Evaluation members in its first meeting on January 31, 2014.

shared with parents, meaning that the test items are not secure, which is a prerequisite for the IE. Moreover, interest in measuring writing skills was expressed by the EQUIP-T Steering Committee and UWEZO does not test writing skills.

The OPM IE team also examined the work of the Southern African Consortium for the Measurement of Education Quality (SACMEQ) approach and instruments, which allows for cross-country comparisons. However, the most recent SACMEQ assessment was conducted in 2007 and SACMEQ tests standard six pupils and EQUIP-T will in focus on the early grades in the first phase. The IE pupil assessment instrument will also be directly based on the Tanzanian primary school curriculum.

Finally, the OPM IE team held discussions with staff from the Research Triangle Institute (RTI) to gain understanding of their work on early grade reading and mathematics in a number of countries and the assessment conducted for the BRN programme in Tanzania. However the purpose of these instruments is different from that of the IE. RTI generously shared their EGRA and EGMA instruments developed for Tanzania with the OPM IE team and these were very helpful in developing the Test Specification Framework for the IE pupil learning instruments.

As described earlier, to develop the IE pupil learning instruments the OPM IE team collaborated closely with a national team consisting of Kiswahili and mathematics specialists from the University of Dar es Salaam, primary school teachers and an expert who had previously developed learning instruments for the UWEZO learning assessment and TWAWEZA Kiufunza project in Tanzania. The Test Specification Framework was used to guide the development of the pupil learning assessments.

Testing the same pupils in both subjects

The IE initially considered testing different pupils in the two subjects: Kiswahili and mathematics, to avoid pupil fatigue and ensure the maximum test administration time per pupil would be 45 minutes given the young age of the pupils. However based on feedback from the Reference Group for the Impact Evaluation and pre-test 1 and pre-test 2 findings, the IE design was revised to test the same standard 3 pupil in both subjects while ensuring a feasible test administration time.

Item difficulty

Given the relatively young age of the pupils and type of learning assessment, the IE administered the pupil tests one-on-one (one enumerator testing one pupil at a time). Following test development in late December 2013, the draft (paper) pupil test instruments were pre-tested to ensure that each item was testing the appropriate curriculum level (standards 1 and 2) and to assess whether the difficulty range of items allows for discrimination for a range of pupil performance. For the latter aspect, the pre-test schools were purposively selected to include both high and low performing schools with respect to education performance. After the first pre-test, analysis was conducted to determine item difficulty, functioning and reliability, and some items were discarded based on this analysis.[13] These revised pupil tests were then pre-tested again (using CAPI) and further revised, and were then pre-tested a third time and revised further during the enumerator training and pilot, before finalisation.

Language

Differences in language spoken at home are likely to influence pupil learning results. The IE seeks to mitigate this issue at two levels. First, as proposed by the Reference Group for the Impact

---

[13] Pre-test 1 report by national team available on request.

Evaluation, it considered language spoken at home when matching the control districts to the treatment districts. Second, the pupil tests include a section on pupil characteristics in which pupils were asked what language they speak at home and this information will be used to control for differences in language spoken at home at the analysis stage.

Poverty measurement

In the technical proposal a pupil tracer to collect poverty data was suggested. However because the pupil learning assessment will be administered one-on-one, which is very time consuming compared to a classroom based test administered to all pupils at the same time, and given a compressed fieldwork window, the pupil tracer was replaced by a module administered at each school. The OPM IE team considered two approaches to collecting these data. First, an asset approach like that used by RTI for the BRN assessment where pupils are asked about their home, for example, "Where do you normally get your water from at home?" and responses are used to construct an asset index. Second, a poverty scorecard approach where parents were asked to come to the school and answer questions about household characteristics such as "How many tables does your household own?" (this approach was proposed by the Reference Group for the Impact Evaluation), and the collected scores are converted to poverty likelihoods. Such a poverty scorecard with a mapping to poverty likelihoods already exists for Tanzania. The final choice was to use the poverty scorecard.

Teacher development needs assessment (TDNA)

Investigation of suitable instruments was guided in part by the possibility that teacher could perceive an assessment covering the pupil curriculum as threatening and partly by the fact that assessments measuring teacher subject knowledge based on the pupil curriculum has not been done in Tanzania in the past. Accordingly, the OPM IE team decided to use a format that is an anonymous, a non-threatening mock pupil test that teachers mark. This type of instrument has been used successfully in other countries, for instance Nigeria. The IE baseline TDNA instrument is based on the format used by a TDNA instrument developed by Dr David Johnson for DFID-Nigeria.

The teacher development needs assessment (TDNA) measures standard 1, 2 and 3 teachers' primary school subject knowledge in Kiswahili and also standards 1-3 and standards 4-7 primary school knowledge of mathematics. Teachers are tested on materials from different curriculum levels to avoid potential floor and ceiling effects.

The TDNA instruments developed for the IE will be used to assess subject knowledge of teachers who teach Kiswahili and mathematics to standards 1-3 and of teachers who teach mathematics to standards 4-7 in 2014, 2016 and 2018. The reason for testing these teachers instead of standard 2 and standard 5 teachers as given in the technical proposal is that the EQUIP-T programme will provide: Kiswahili INSET for standards 1-3 curriculum levels to standards 1-3 teachers (phase 1); mathematics INSET for standards 1-3 curriculum levels to standards 1-3 teachers (phase 1); and mathematics INSET for standards 4-7 curriculum levels to standards 4-7 teachers (phase 2).

As described above, the OPM IE team worked with the same national team that developed the pupil tests to develop the TDNA instruments. This work was also guided by the Test Specification Framework designed by the OPM IE team and the instruments were revised based on the pre-test findings.

Item difficulty

The TDNA instruments were pre-tested to ensure that the instruments cover the whole primary school curriculum (only a minimum number of items for the standards 1 and 2 curriculum levels were included as these items are very basic), and that the difficulty range of items allows for discrimination for a range of teacher performance to avoid potential ceiling and floor effects. For the latter aspect, the pre-test schools were purposively selected to include both high and low performing schools with respect to education performance. After the first pre-test, analysis was conducted to analyse item difficulty, functioning and reliability, and some items were replaced based on the analysis. The revised TDNA instruments were pre-tested a second time using CAPI, and then further revised, and were then pre-tested a third time, and then further revised during the enumerator training, before finalisation.

Lesson observation

To develop this instrument the OPM IE team reviewed existing lesson observation instruments including the work done in Tanzania by UNICEF (Hardman and Dachi 2012). There are three choices which have been shown to work well in the Tanzanian context: (i) frequency observation; (ii) timeline analysis; (iii) computerised observation schedule (gives frequency and duration information). Based on observations from the school visits during the November 2013 scoping mission and pre-test 1, the OPM IE team chose the frequency observation option. The lesson observation instrument worked very well during all three pre-tests.

The lesson observation instrument of standard two Kiswahili and mathematics lessons captures teacher pedagogy, inclusive behaviour of teachers with respect to gender; key teacher behaviours (drawn from research on effective teaching and from pedagogical areas the EQUIP-T programme is seeking to change); and availability of lesson plans. Key skills observed include: (i) introducing lessons; (ii) questioning pupils; (ii) giving feedback to pupils; (iv) using instructional materials; (v) building rapport with pupils; and (vi) concluding lessons. The lesson observations was also used to capture whether teacher interaction with pupils in the classroom is gender balanced.

Teacher interview

The interviews with standard one to three teachers (initial focus of EQUIP-T) capture: background information (gender, age, years of teaching, qualifications), self-assessment of job satisfaction, frequency/nature of in-service training received, frequency/nature of lesson observation by head teacher/others and nature of feedback, and frequency/nature of performance appraisal by head teachers.

EQUIP-T will seek to raise teacher motivation by developing a Teacher Morale Toolkit. To measure (teacher) motivation in a meaningful and appropriate way is very difficult. One option is to use subjective questions to teachers such as "I am fully satisfied with my current job" with response categories of the Likert-scale type: "strongly agree", "agree", "disagree" and "strongly disagree". Several questions of this type were tested during the pre-tests but did not perform well and there was very limited variation in the responses. These questions were therefore not used for the IE. The OPM IE team shared this information to the MA when sharing the IE instruments in February 2014, to help inform the development of instruments for the EQUIP-T baseline monitoring survey. Another approach is a self-assessment of jobs satisfaction based on a visual in the form of a ladder with ten steps, to help concretize the question, which was used for the IE baseline survey.

Head teacher interview and school records checks

The interviews with head teachers and data collection from school records capture: background information on head teachers (gender, age, years of experience, qualifications), school background information (enrolment, teachers, physical facilities, school timetable, number of days school open

etc.), teacher attendance (by records and by headcount on the day), pupil attendance (by records and by headcount on the day), and frequency/nature of school planning/management in-service training received.

Instructional hours, a measure of instructional time, is measured using a combination of information from the different instruments. The following approach was used to capture this: first, collection of data on number of days a school is open in a year; second collection of data on timetabled lesson minutes by subject per day; and third, collection of data on whether a teacher is timetabled to teach or not and is in the classroom as timetabled. This was measured separately for Kiswahili and for mathematics.

3.4   Impact indicators

This section presents the quantitative indicators to be used to measure EQUIP-T programme impact in 2016 and 2018. For a list and definitions of all the IE quantitative indicators presented in Part B Results see Annex F.

The IE will quantitatively measure impact on pupil learning outcomes and selected outputs for EQUIP-T component 1: Teacher capacity and professional performance and for EQUIP-T component 2: School leadership and management. The impact indicators are listed below by expected EQUIP-T programme outcome and outputs.

EQUIP-T outcome: Improved quality of education, especially for girls

- **Indicator 1a**: Proportion of pupils in the bottom Kiswahili performance band, disaggregated by gender (%)
- **Indicator 1b**: Proportion of pupils in the top Kiswahili performance band, disaggregated by gender (%)
- **Indicator 1c**: Proportion of pupils in the bottom mathematics performance band, disaggregated by gender (%)
- **Indicator 1d**: Proportion of pupils in the top mathematics performance band, disaggregated by gender (%)

This will allow 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.

EQUIP-T output 1: Enhanced professional capacity and performance of teachers

- **Indicator 2a**: Mean TDNA Kiswahili score for standards 1-3 teachers (% correct items on TDNA Kiswahili)
- **Indicator 2b**: Mean TDNA mathematics score for standards 1-3 teachers (% correct items on TDNA mathematics)
- **Indicator 2c**: Mean TDNA mathematics score for standards 4-7 teachers (% correct items on TDNA mathematics)
- **Indicator 3**: Proportion of standards 1-3 teachers that could <u>not</u> show any evidence of assessing pupil academic progress (%)

Teachers who teach standards 4-7 mathematics will receive relatively less exposure to the EQUIP-T programme as activities aimed at this group will start later than those aimed at teachers who

teach standards 1-3 mathematics. Therefore it will, other things remaining equal, be more difficult to detect any change in teacher subject knowledge attributable to the programme for this group.

TDNA results for teachers of standards 1-3 and standards 4-7 mathematics may also be combined into one indicator for all mathematics teachers to potentially increase the sample size and minimum detectable effect size.

EQUIP-T output 2: Enhanced school leadership and management skills

- **Indicator 4a**: Proportion of standard 1-3 teachers who report lesson observations by head teacher in last 30 days (%)

- **Indicator 4b**: Proportion of standard 1-3 teachers who report head teacher feedback on lesson plans in the last 30 days (%)

- **Indicator 4c**: Proportion of standard 1-3 teachers who report participation in performance appraisal (%) (TBC)

- **Indicator 5a**: Teacher school absenteeism (%)

- **Indicator 5b**: Teacher classroom absenteeism (%)

The IE impact indicators listed above will be measured at three points in time: 2014 (baseline), 2016 (midline) and 2018 (endline).

These indicators sit at two different levels: (i) pupil (indicators 1a-1d) and (ii) teacher (indicators 2a-2c, 4a-c, 3 and 5a-b based on different teacher samples), and will be used to measure change in selected programme outcomes and outputs over time.

The power of detection (minimum detectable change) for these indicators based on the theoretical power calculations in OPM (2014a) is anticipated to be sufficient to detect any such changes. Nevertheless, this will not be confirmed until the empirical power calculations are done in 2016 (see section 3.2.4). The teacher samples for impact indicators 2a-2c, 3 and 4a-4c are relatively small. This limitation was beyond the IE design given the school sample size chosen by DFID and was set out in the inception report (OPM 2014a).

3.5   Baseline estimates and impact estimation approach

As described in section 3.1.1, the IE will estimate the impact of the EQUIP-T programme as a whole and for this will use the impact indicators outlined in section 3.4.

Section 3.5.1 presents baseline estimates for the impact indicators and key characteristics using unweighted data for the treatment and control groups.[14] This means that the treatment estimates in this section are *not* comparable to the treatment estimates contained in Volume I as those are weighted to be representative of the programme treatment area. Next, section 3.5.2 discusses the approach to estimating impact in 2016 (2018) using matched samples, and provides an illustration of how PSM can be used to generate matched samples that are balanced.

3.5.1   Baseline estimates

The options for creating a counterfactual were restricted given that the EQUIP-T programme districts were pre-selected by the programme (see section 3.1), and the group of eligible potential

---

[14] Unweighted estimates are shown in this section because the control group was selected using PSM instead of being randomly sampled and therefore there are no survey weights for the control units.

control districts was limited (see Box 1). Under these circumstances a PSM approach was the best option available to select the control group as described in the EQUIP-T Impact Evaluation Inception and Baseline Design Report (OPM 2014a). However, given that treatment and control groups were not randomly selected from the same overall population the control and treatment groups may still exhibit some statistically significant differences in particular indicators and characteristics.

Table 3 presents unweighted mean estimates for the impact indicators and differences in group means, P-values from the testing of whether the differences in group means are statistically significantly different, and sample sizes. The treatment estimates presented in this section are different from those in Volume I as they are unweighted as discussed above.

The estimates in this section are for the unmatched samples, that is, PSM has not yet been used to generate matched baseline samples, which means that there may still be some significant differences across the control and treatment groups (Dehejia and Wahba 2002). This matching will be done in 2016 for the impact estimation (see section 3.5.2 and Box 2).

In Table 3, about half of the impact indicators are not significantly different for the treatment and control groups for the unmatched samples, with pupil mathematics scores, teacher TDNA mathematics scores, and all but one of the SLM impact indicators being statistically similar across the two groups. The treatment and control groups are significantly different with respect to the pupil Kiswahili test results (impact indicators 1a and 1b), teacher Kiswahili subject knowledge (impact indicator 2a), the proportion of teachers measuring pupil academic progress (impact indicator 3) and teacher school absenteeism (impact indicator 5a).

One plausible reason for the significant differences in the pupil test scores across the control and treatment groups is language spoken at home, but not only whether a pupil speaks Kiswahili or a local language other than Kiswahili at home, but how 'well' a pupil speaks Kiswahili. To address this issue the IE midline survey will collect additional information on language such as language spoken by pupils (not only if they speak Kiswahili at home or another local language), how many different languages they speak and language spoken by mother and by father. Language spoken also tends to be correlated with poverty in this context, and therefore the midline survey will also collect information on whether pupils need to work for pay and how much they help out at home, which would affect time available for studying outside of school. These variables will be used for the matching model to balance the sample (see Box 2).

Table 3 Baseline impact indicator estimates

Impact indicators		Treatment	Control	Difference	P-value	N
EQUIP-T outcome: Improved quality of education, especially for girls						
1a	Proportion of pupils in the bottom Kiswahili performance band (%)	39.3	24.3	14.9***	0.000	2,892
	Proportion of boys in the bottom Kiswahili performance band (%)	37.6	23.5	14.1***	0.000	1,500
	Proportion of girls in the bottom Kiswahili performance band (%)	42.3	25.9	16.4***	0.000	1,486
1b	Proportion of pupils in the top Kiswahili performance band (%)	12.3	16.7	-4.4***	0.001	2,892

	Proportion of <u>boys</u> in the top Kiswahili performance band (%)	11.4	13.4	-2.0	0.245	1,500
	Proportion of <u>girls</u> in the top Kiswahili performance band (%)	12.5	19.6	-7.0***	0.000	1,486
1c	Proportion of pupils in the bottom mathematics performance band (%)	7.5	6.6	1.0	0.314	2,880
	Proportion of <u>boys</u> in the bottom mathematics performance band (%)	6.0	5.3	0.7	0.552	1,497
	Proportion of <u>girls</u> in the bottom mathematics performance band (%)	9.2	8.2	1.1	0.474	1,476
1d	Proportion of pupils in the top mathematics performance band (%)	6.1	6.2	-0.2	0.861	2,880
	Proportion of <u>boys</u> in the top mathematics performance band (%)	8.2	9.3	-1.1	0.473	1,497
	Proportion of <u>girls</u> in the top mathematics performance band (%)	3.8	3.3	0.5	0.582	1,476
EQUIP-T output 1: Enhanced professional capacity and performance of teachers						
2a	Mean TDNA Kiswahili score for standards 1-3 teachers (% correct items on TDNA Kiswahili)	58.9	61.0	-2.1*	0.072	510
2b	Mean TDNA mathematics score for standards 1-3 teachers (% correct items on TDNA mathematics)	51.7	51.5	0.2	0.910	505
2c	Mean TDNA mathematics score for standards 4-7 teachers (% correct items on TDNA mathematics)	66.4	68.6	-2.2	0.151	562
2b & 2c	Mean TDNA mathematics score for standards 1-7 teachers (% correct items on TDNA mathematics)[2]	59.4	60.3	-1.0	0.481	1,023
3	Proportion of standards 1-3 teachers that could not show any evidence of measuring pupil academic progress (%)	30.1	23.0	7.1**	0.037	681
EQUIP-T output 2: Enhanced school leadership and management skills						
4a	Proportion of standard 1-3 teachers who report lesson observations by head teacher in last 30 days (%)	49.2	47.4	1.8	0.639	677
4b	Proportion of standard 1-3 teachers who report head teacher feedback on lesson plans in the last 30 days (%)	90.9	92.3	-1.4	0.504	680
4c	Proportion of standard 1-3 teachers who report participation in performance appraisal (%)	27.7	25.7	2.0	0.568	679
5a	Teacher school absenteeism (%)	13.1	8.9	4.2***	0.003	1,936

5b	Teacher classroom absenteeism (%)	67.8	70.1	-2.3	0.349	1,440

Notes: 1) Group means; 2) Some teachers report teaching both standards 1-3 and standards 4-7. Asterisks indicate statistical significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Background characteristics for the treatment and control units are shown in Table 4. The proportions of boys, poorer pupils and pupils who speak Kiswahili at home are significantly smaller in the treatment area than in the control area, and the mean pupil age is significantly different. The midline survey will collect additional pupil background information including the ones mentioned above on language, child labour, and chores. These variables will be used to build a matching model (see Box 2), to balance the pupil sample.

For teachers who teach standards 1-3, the only significant differences across the treatment and control areas are that the proportion with a Form 4 academic qualification is larger and the proportion of teachers with a primary education is smaller in the control area than in the treatment area. On the other characteristics such as age, years of experience and professional qualifications there are no significant differences across the two groups.

Head teacher and school characteristics are also shown and are similar across the treatment and control areas with no significant differences. However, it should be noted that this sample is relatively small consisting of 200 head teachers/schools with potential implications for statistical inference.

Table 4 Baseline background characteristics

Background characteristics	Treatment	Control	Difference	P-value	N
Pupils[3]					
Girls (%)	51.8	47.7	4.1**	0.027	2,987
Age (years)	10.3	10.2	0.2***	0.003	2,519
Speak local language other than Kiswahili at home (%)	78.0	75.2	2.7*	0.079	2,987
From household (predicted to) be below national poverty level (%)	33.3	39.5	-6.1***	0.001	2,893
Teachers (standards 1-3)					
Female (%)	59.9	58.2	1.6	0.664	681
Experience working as a teacher (years)	15.2	14.2	1.0	0.335	681
Primary education (%)	12.5	7.1	5.4**	0.019	681
Form 4 (%)	77.5	85.5	-8.0***	0.007	681
Form 6 (%)	3.3	1.4	1.9	0.103	681
Academic certificate (%)	6.4	6.0	0.4	0.822	681
Professional certificate (%)	95.7	96.0	-0.3	0.849	680
Professional diploma (%)	1.8	2.0	-0.2	0.880	680
No professional qualification (%)	0.3	1.1	-0.8	0.196	680
Received in-service training 2012-13 (%)	7.6	4.8	2.8	0.136	681
Head teacher					
Female (%)	16.0	19.0	-3.0	0.579	200
Time as head teacher at this school (years)	4.2	3.8	0.4	0.494	200
Primary education (%)	0.0	0.0	0.0	N.A.	200

Form 4 (%)	78.0	80.0	-2.0	0.730	200
Form 6 (%)	13.0	12.0	1.0	0.832	200
Academic certificate (%)	9.0	8.0	1.0	0.801	200
Professional certificate (%)	89.0	85.0	4.0	0.403	200
Professional diploma (%)	10.0	12.0	-2.0	0.653	200
Received in-service training 2012-13 (%)	9.8	14.6	-4.8	0.316	188
School					
Total enrolment standards 1-7	613.8	560.9	52.8	0.230	200
Pupil-teacher ratio (PTR)	59.8	57.2	2.6	0.455	200
Total number of days school was open in 2013	191.8	191.8	0.0	0.989	181
School has school management committee and minutes are available from its last meeting (%)	91.0	94.0	-3.0	0.423	200
School holds parent-teacher meeting more than once a year and minutes from last meeting available (%)	67.0	74.0	-7.0	0.280	200
Number of ward education coordinator (WEC) visits in 2013 (%)	6.6	6.1	0.5	0.401	197
Schools visited by district school inspector (DSI) in 2013 (%)	61.0	68.0	-7.0	0.303	200
Schools with functional toilet (%)	96.0	91.0	5.0	0.153	200
Schools with drinking water available (%)	38.0	36.0	2.0	0.771	200

Notes: 1) Group means; 2) Artificial weights were constructed for the control group observations as these were not randomly sampled; 3) These are the estimates for the full pupil sample. Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.

3.5.2   Impact estimation approach

A PSM-DID technique will be used to estimate impact in 2016 to allow for time-varying selection bias due to differences in initial observable characteristics (see above), and selection bias due to unobservable time-invariant characteristics.

The PSM-DID technique estimates the impact of the EQUIP-T programme based on the change in outcomes compared between treatment and control groups, taking into account differences between the treatment and control groups that are due to imperfect matching only to the extent that these are observed in the data or do not change during the life of the programme. Differences between the two groups which are not observed in the data collected, and which change during the life of the programme, would violate the common trends assumption, and constitute the main risk of bias.

The IE will estimate the average treatment effect on the treated (ATT) for the treated units and *matched* control units (see Box 2 for an illustration) that are on the common support (Khandker et al. 2010). This will be estimated as the difference between the change in the outcome/outputs (Y) of interest over time for the treated group ($Y_{i2}^T - Y_{i1}^T$) and for the control group ($Y_{j2}^C - Y_{j1}^C$):

$$ATT = (Y_{i2016}^T - Y_{i2014}^T) - \sum_{j \in C} \omega(i,j) (Y_{j2016}^C - Y_{j2014}^C)$$

Where T indicates the treated group, C indicates the control group, and $\omega(i,j)$ is the weight (using PSM) assigned to control unit j matched to treatment unit i.

At midline, PSM will be used to balance the samples by matching (Dehejia and Wahba 2002, Khandker et al. 2010) and then combined with DID to estimate impact based on the matched samples. Box 2 illustrates how PSM can be used to achieve a balanced sample. It uses a sub-sample of the pupils who took the Kiswahili test and for whom age data is available (this is a different (unmatched) pupil sample to that in Table 3 so estimates should not be compared).

Box 2 Illustration of how matching will be used to achieve balanced samples

Note: The baseline estimates in this box should <u>not</u> be compared to other estimates in this report, and are included for illustrative purposes only.

For PSM, the first step is to estimate the propensity score. For the pupil sub-sample, this was done using a logit model that includes covariates correlated with treatment status and the outcome of interest and that will not be influenced by the programme. For this illustration, the covariates included the programme selection criteria; district characteristics capturing poverty and education performance; school characteristics including the PSLE pass rate, infrastructure and language of instruction; and pupil characteristics in the form of age, whether the pupils speak Kiswahili or another local language at home, household size, and a selection of asset variables.

For this illustration a kernel matching estimator was used that compares the outcome of each treated unit to a weighted average of the outcomes of all the control units, assigning the largest weight to the control units with the propensity scores closest to the treated unit (Heinrich et al. 2010). Using a kernel estimator poor matches may be included so to address this issue all observations that were not on the common support were excluded[15].

Table 5 presents (unweighted) mean estimates for 2,283 treatment and control pupils who reported their age, differences in the group means, P-values for the test of significant differences in group means, and sample size for the unmatched and matched samples.

This shows that in the unmatched sample there are statistically significant differences in the proportion of pupils in the bottom and top bands of the Kiswahili performance bands respectively, but in the matched sample these differences are not significant. In the unmatched sample 34.7% of pupils that will participate in the programme are in the bottom performance band and 20.9% of the pupils who will not participate. This compares to 34.2% of pupils that will participate in the programme and 30.0% of non-participating pupils respectively in the matched sample. This illustrates how PSM will be used achieve a balanced sample in 2016.

Table 5 Unmatched and matched sample estimates for a sub-sample of pupils

Impact indicator		Unmatched sample				Matched sample				
		Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value	N
1a	Proportion of pupils in the bottom Kiswahili performance band (%)	34.7	20.9	13.8***	0.000	34.2	30.0	4.2	0.369	2283
1b	Proportion of pupils in the top Kiswahili performance band (%)	13.7	18.5	-4.8***	0.001	13.8	13.0	0.8	0.636	2283
Notes: 1) Unweighted estimates; 2) Bootstrapped standard errors.										

---

[15] Choice of and sensitivity analysis for bandwidth and trimming are not discussed here, but would be conducted as relevant for the impact estimation in 2016.

3.6 Baseline and follow-up surveys

The time between the baseline and midline surveys is set at two years and between the baseline and endline at four years. If the programme works as intended, two years would likely be a long enough period to detect impact on intermediate outcomes such as teacher capacity and performance and some school leadership and management behaviours, but two years is a relatively short period to expect any detectable impact on pupil learning outcomes, which would have to work its way through changes in the intermediate outcomes. However, at the time of the endline survey in 2018, it would be reasonable to expect to be able to detect changes in pupil learning levels if the programme has worked.

The baseline quantitative survey took place from the end of March to mid-May 2014. The follow-up surveys in 2016 and 2018 will take place at approximately the same time of year to ensure that tested pupils have been exposed to the approximately the same amount of teaching.

3.7 Possible risks and limitations of quantitative component

Protecting the impact evaluation design over time

Concern for the methodological soundness of the IE design does not end with the initial design and identification of treatment and control groups. As the IE is a continuous process of observing any changes in school and pupil performance attributable to the programme over time, the IE design needs to be protected for the duration of the impact evaluation.

The most common risk in longitudinal surveys is potential contamination of the selected impact evaluation areas by third party interventions that may affect the outcomes and outcomes of interest for the evaluation of the programme. To address this risk the IE will monitor treatment and control districts for any new education programmes or projects that could potentially confound EQUIP-T impact and will be in regular contact with the MA, DFID and donors in education to try to obtain such information. However, having access to this type of information is only part of the process; ultimately the Government determines which districts in Tanzania would receive any new education programme or project. Therefore the IE will liaise with the MoEVT and PMO-RALG to explore the possibility of any potential new education programme to be rolled out in the IE districts. Moreover, the IE team will discuss this potential risk with DFID and the option of DFID raising it with the Education Development Partners Group to help minimise any potential contamination by other education programmes.

One particular risk at this stage is the roll out of another large education programme, LANES funded by the Global Partnership for Education that will seek to achieve similar outcomes to the EQUIP-T programme over the period 2014-2017. LANES is in the early stages of planning its interventions which seem, at this stage, to be somewhat similar in nature to part of the teacher component of EQUIP-T, although the delivery model is different.[16] The LANES programme will target a selected number of early grade teachers for training and support in teaching children basic numeracy and literacy skills (under the newly emerging national 3Rs approach). The intention of the LANES programme is to cover all districts which are not already covered by an early grade support programme (so this will exclude EQUIP-T programme districts). Clearly if the IE control districts are included in LANES, and the intervention works, this will confound the EQUIP-T impact estimates of pupil learning.

---

[16] The EQUIP-T model of INSET training and support is largely school-based, while the LANES model is a central cascade model (based on information available in December 2014).

To minimise the risk of contamination from LANES, the OPM IE team is communicating with DFID on this issue, and also held discussions in November 2014 with SIDA who are the supervising entity for LANES. This issue also was also raised and discussed with senior government representatives at the EQUIP-T IE Reference Group meeting in November 2014. The first stage mitigation strategy is to seek the optimal sequencing of LANES if there is a phased roll-out, such that the teachers in the eight control districts participate in the last phase of training, hopefully after the mid-line research for the IE (2016). In addition, the IE team will request regular updates on LANES implementation, as well as any monitoring information on its performance. Prior to the design of the IE mid-line research, the risk of contamination by LANES will be reviewed as it may be necessary to substantially alter the design of the IE in order to meet its objectives given the evolving context. The most obvious possibility, given the work already done on expanding and analysing the programme TOC, would be use a 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.

Confounding TOC failure and implementation failure

Another risk is the possibility of confounding 'theory of change failure' with 'implementation failure'. That is, the EQUIP-T programme may not have the expected impact because the TOC underlying the EQUIP-T programme design (see Cambridge Education 2014a, OPM 2014a) is not correct in its assumptions about causal linkages, although implemented according to programme design (TOC failure). Alternatively, the programme may not achieve the expected impact despite a valid TOC because of implementation deviating from the design. To address the risk of TOC failure the IE will at the baseline, primarily through the qualitative component assess the TOC hypotheses and assumptions (section 294.3). For the second type of risk, implementation failure, the impact evaluation will monitor EQUIP-T implementation based on data from the MA's MIS.

Adequate and timely programme implementation data

Finally, to ensure correct measurement of EQUIP-T impact, the IE team will require timely information on the programme implementation rate (contents, intensity and timing) from the MA for each of the four EQUIP-T components covered by the IE. At the baseline for example, the OPM IE team will need information from the MA, initially on the planned implementation rate, and subsequently on the actual implementation rate on an on-going basis. For example, for component 1 on teacher performance, this would include the number of teachers who received INSET, the type of INSET and timing etc. for all EQUIP-T districts and schools. Points at which this information will be required from the MA and agreements on this are shown in the EQUIP-T Impact Evaluation Inception Report (OPM 2014a). If the MA cannot provide this information as specified, the IE will not be able to produce the deliverables as scheduled.

Other limitations of the quantitative component

Table 6 lists some other limitations of the quantitative component and mitigation factors. As well as these more general limitations, some specific limitations to the baseline analysis arose during the survey. One issue was the extent of non-response, which was generally low but in some areas was higher than would be desirable. Annex H contains statistical tables which show intended and actual sample sizes for all of the key indicators reported in Volume I. The rate of non-response was

particularly high for the self-reporting of the age of pupils. About 16% of age data is missing for the tested standard three pupils. In hindsight, it would have been sensible to also collect this data from their parents and this will be done in the next round. Another area where there was lots of missing data was on capitation grants. For some of the indicators in this area, about one-third of data is missing. In the Volume I report, the discussion on capitation grants includes various caveats about the data, and after careful analysis some of the data was rejected as unusable. Generally the collection of capitation grant information was problematic in many cases. The financial records for the capitation grant are complex, head teachers, especially those new to the job don't understand them well, and enumerators complained that it was sometimes difficult to work out the timing of various payments, for example. One other area where sample sizes are less than expected, but (largely) not due to non-response, is the TDNA that was given to teachers. Here the samples are lower than the anticipated simply because there were not enough teachers in the school to administer the instruments to.

Table 6 Possible limitations of quantitative component

Possible limitation	Why is this limiting and mitigating factors
Not possible to change survey instruments after the baseline	If there are changes to the EQUIP-T programme design after the baseline the instruments will not be able to measure this. The indicators included have been carefully considered to ensure they capture key EQUIP-T outcome and outputs that are likely to remain unchanged.
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
Language spoken at home	Pupils that speak do not speak Kiswahili at home may be at a disadvantage for the pupil testing. Prevalence of ethnic languages will be used as one of the matching criteria at the district level and for school matching. The pupil tests included questions on language spoken at home and this information will be used to control for differences in language spoken at home at the analysis stage.
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.
Relative short time period for assessing EQUIP-T impact on pupil learning	The first EQUIP-T activities were rolled out in July 2014 and the midline impact evaluation survey will take place in March-May 2016 but two years is a relatively short time to expect any EQUIP-T impact on pupil learning. It may only be possible to demonstrate any such impact after four years (in 2018). The baseline survey will also measure a series of intermediate programme effects, such as teacher capacity and performance, which would be expected to change more quickly assuming the EQUIP-T programme works as intended.

4 Qualitative research design

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, fieldwork, analysis, structured and unstructured methodologies, using theory to inform generalizability and development of a strengthened TOC. A more detailed description of the qualitative sampling, tools and fieldwork can be found in Annex J.

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. The specific objectives of the qualitative component are to:

- Identify and test priority hypotheses in the EQUIP-T TOC in terms of the EQUIP-T programme design, implementation and context, at district, ward, school and community levels. Such a theory-based assessment (although partial) approach will provide some indication of the likely impact of the EQUIP-T programme if scaled up;

- Provide an analysis of the perceptions of key stakeholders of the relative contributions of the different EQUIP-T components to the outcomes of interest at district, ward, school, and community level; and

- Investigate and explore the mechanisms behind hypotheses or unexpected findings from the quantitative survey.

Similarly to the quantitative research, the qualitative research collects data at three points in time: 2014, 2016 and 2018.

The qualitative methodology is informed by the objectives of the qualitative research: to develop and test 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 quantitative findings. This range of objectives implies a qualitative methodology that is both confirmatory (testing existing hypotheses) and exploratory (explaining impacts, developing new hypotheses and capturing unexpected impacts).

4.2   Rigour

A major methodological 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). How the qualitative research design addressed these issues, in terms of sampling, fieldwork, and analysis is discussed in the relevant sections below.

4.3   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, the impact evaluation uses the TOC for this purpose (Cartwright and Hardie 2012).

The EQUIP-T TOC informs the impact evaluation as a whole, but is particularly important for the qualitative component of the research because it permits stronger generalisation and attribution of impact. Specifically, 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 impact evaluation will tell us whether impact can be attributed to EQUIP-T. But the impact evaluation also uses (primarily) qualitative data 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 Annex G) and then setting specific questions for (more) structured instruments and observations. This will allow some consideration of the likely impact of EQUIP-T. As set out in the EQUIP-T Impact Evaluation Inception Report (OPM 2014a), this generalisation is not perfect, because it cannot account fully for the existence of other education programmes in these districts, but the process of generalising takes these programmes into account as context, based on available secondary documentation on these programmes.

The baseline quantitative and qualitative fieldwork assists in further focussing subsequent rounds of qualitative data collection on specific issues in the context and on causal pathways that are considered by stakeholders and researchers to be most significant in terms of either their contribution to overall impact or their uncertainty. To this extent, the evaluation is based in theory, and therefore amenable to a certain amount of evidence about what would happen if different components of EQUIP-T were scaled up.

As explained above, the evaluation design is based on the quantitative research estimating impact and outcomes and the qualitative research exploring causal explanations for positive outcomes, to provide a rigorous assessment of the EQUIP-T programme in the districts in which it is being implemented.

4.3.1   How will the programme theory of change be strengthened?

The EQUIP-T programme TOC and the EQUIP-T programme design (Cambridge Education 2014b), which includes a set of hypotheses and assumptions, has provided the starting point to develop an expanded TOC. The EQUIP-T TOC, while appropriate for programme management purposes, is not currently sufficiently developed to generate a comprehensive set of hypotheses about causal pathways and context.[17]

From this starting point, the TOC development was undertaken in two stages. First, as part of the inception phase, relevant EQUIP-T documentation and wider education literature were reviewed (Stephens 2013), and interviews were conducted with key stakeholders in education in Tanzania, including EQUIP-T programme staff. This led to the development of a more specific and elaborated qualitative evaluation matrix (Annex G). In this matrix, the current TOC is adapted for each component of the EQUIP-T programme by focusing on the specific constraints that actors at the district, community and school levels face. While children and their ability to learn to their full potential are at the core of the overall EQUIP-T TOC, teachers, head teachers, district officials and communities are key to the capability constraint sub-theories of change guiding the individual components of the EQUIP-T programme. This matrix forms the basis for the initial set of confirmatory hypotheses and structured questions employed in the baseline qualitative research. A further review of evidence regarding each individual assumption in the qualitative evaluation matrix was undertaken in order to allow the researchers to compare the evidence against the contextual data collected during the baseline qualitative research.

Second, as part of the qualitative and quantitative baseline fieldwork, a set of contextual factors was investigated and areas to help the impact evaluation team develop an enhanced TOC for the impact evaluation identified. This enhanced TOC will set out the key causal pathways and

---

[17] This limitation was noted by the SEQAS March 7, 2014.

contextual assumptions in more detail, and will allow for the identification, and testing through subsequent rounds of fieldwork, of more specific hypotheses around EQUIP-T's impact.

4.4   Sampling

Nine schools across three districts were sampled as sites for the qualitative research. The following section will briefly outline the sampling approach of the qualitative research. More technical detail on how the sampling was conducted can be found in Annex J.

4.4.1   Addressing rigour through sampling

Methodological rigour in qualitative research is not best established through a statistically representative sample, because results cannot be quantified and aggregated in the same way as quantitative data. Nonetheless, as in quantitative research, rigour in qualitative research can be achieved through "systematic and self-conscious research design, data collection, interpretation and communication (Mays and Pope 1995, p110)."

As with most qualitative research, the chosen approach to sampling for this IE is theoretically informed and designed to generate responses from small numbers of individuals and groups that are representative (though not statistically) of groups relevant to EQUIP-T, and which allow some identification of heterogeneous impact. Specifically, purposive 'typical case sampling' (at the highest level, sampling an average district in terms of education outcomes) and 'extreme case sampling' (sampling high and low performing districts in terms of selected education inputs, outputs and outcomes) was used. This is not designed to produce results that are generalizable in the same sense as quantitative data. Rather, the generalizability of the qualitative research results derives from the extent to which they are embedded in a TOC that has some validity in a wider context. This form of sampling allows exploration of what EQUIP-T is doing in a typical case, but also performance in 'good' and 'bad' cases, which helps explain the reasons behind success and failure.

4.4.2   Selection of districts

The quantitative survey sampling frame was also used to select districts for the qualitative research. Control group districts were removed as possible sites as the qualitative research is primarily concerned with the changes within programme districts.

Three districts were selected using purposive 'typical case sampling' and 'extreme case sampling' on education inputs, outputs and outcomes at the district level, with consideration given to the social and economic resources of the population in each district. An analysis of socio-economic resources, educational resources and the average academic achievement of pupils within schools was undertaken using UWEZO data and primary school leaving examination (PSLE) results to select three districts.

The final district selection was based on the following criteria:

- One district from each of the above categories must be represented, in line with the qualitative research sampling design, requiring typical and extreme case sampling (high, typical and low performing districts);
- Pupil performance (using UZEWO scores) must represent three of the four performance quartiles at the district level, in order to ensure that the typical and extreme cases cover absolute performance, not only expected performance based on resources;

- The level of pupil and school social and economic resourcing should be diverse in order to provide the highest chances of heterogeneous impact;

- The three districts must be taken from different Regions in order to provide as much coverage of diver context across the treatment districts as possible; and

- Travel time between the three locations should be practical and represent Value for Money, given the time and resources constraints on the evaluation.

Given the importance of gender to EQUIP-T, the selected districts were then reviewed in light of lower secondary enrolment by gender. Two of the selected districts had higher than average female enrolment in lower secondary education and one had significantly lower enrolment.

4.4.3   Selection of schools

Three schools within each of the three selected districts were selected using purposive 'typical case sampling' and 'extreme case sampling'. Schools were selected in order to ensure that they did not only represent diversity in relation to relative performance (as compared to other schools in the district), but also to ensure diversity in relation to the rate of improvement. The measure is indicative as it is open to cohort effects because the average GPAs only take two years of result into consideration. However, an indicative improvement variable is important as the more diverse the school case studies, the more likely that heterogeneous impacts of EQUIP-T will be identified. Consideration was also given to the pupil-teacher ratio and the proportion of teachers who are qualified.

Table 7 Qualitative school case study sites in Mpwapwa, Uyui and Kishapu

Region	District	School	Performance	Academic improvement	Gender Parity
Dodoma	Mpwapwa	Mpwapwa School A	Performing typically	Medium to high	Not close to gender parity
	Mpwapwa	Mpwapwa School B	High performing	High	Almost achieving gender parity
	Mpwapwa	Mpwapwa School C	Not performing well	Low	Not close to Gender parity
Tabora	Uyui	Uyui School A	Typical performance	Above average	Not close to gender parity
	Uyui	Uyui School B	Not performing well	Low	Closer to gender parity
	Uyui	Uyui School C	High performing	High	Closer to gender parity
Shinyanga	Kishapu	Kishapu School A	Typical performance	Low	N/A[1]
	Kishapu	Kishapu School B	Good Performance	High	N/A
	Kishapu	Kishapu School C	Not performing well	High	N/A

Note: Gender parity data was not available for Kishapu schools

4.5   Semi-structured interview instruments and discussion guides

4.5.1   Structured and unstructured methodologies

The qualitative part of the IE makes use of two research techniques or 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 testing of pre-specified

hypotheses and unstructured methodologies allow for unanticipated or context specifics to be captured and for new hypothesis to be developed.

Figure 2 outlines which types of techniques were used to collect information from different participants in the qualitative research. Both the KIIs and the FGDs were designed to explore themes and questions that relate to more than one EQUIP-T component. In other words, the purpose of the different KIIs and FGDs differs insofar as it relates to different areas of interest and different EQUIP-T interventions.

Figure 2 Instruments administered by participant group

Instrument	Head teachers	Pupils	School committee	Parents	Teachers	Community Leader	EQUIP-T staff	District officials	Ward officials	Regional officials
FGDs		X	X	X	X					
KIIs	X					X	X	X	X	X

4.5.2   Key informant interviews and focus group discussion tools

The tools for the key informant interviews and focus group discussions were developed in order to capture information on the core areas to probe outlined in the evaluation matrix - derived from the underlying assumptions in the TOC - and to further investigate unexpected quantitative findings. Each of the instruments focused on the core areas to probe relevant to the group or individual participating in the research. The comprehensive evaluation matrix guiding the qualitative research can be found in Annex G. Table 8 below provides an overview of the areas investigated by component and qualitative research tool.

The KIIs were semi-structured and organised around the different levels of constraints outlined in the EQUIP-T programme TOC. The instruments for all KIIs included an icebreaking discussion and then explored the core areas to probe across the four programme components, as outlined in the evaluation matrix extract below. These discussions were structured, starting with local, school, and district and regional level constraints. This approach provided a degree of standardisation whilst at the same time allowing the impact evaluation team's qualitative researchers enough flexibility to pick up on interesting themes, topics and concerns as they emerged.

The FGDs included icebreaking discussions tailored to the group and then explored the core areas to probe across the four programme components. In addition, the tools also built on selected quantitative findings and explored them further during discussion. Participatory methods were used throughout the FGDs and included problem tree analysis, "a day in the life of", force-field analysis, and the use of vignettes (see Annex J for details on the instruments and participatory methods used).

Table 8 Core areas to probe and sources of information from the qualitative evaluation matrix

Area of impact	Core areas to probe (including change over time for each)	Source of information

Output 1: Enhanced professional capacity and performance of teachers	- How the INSET teacher training is implemented in the region - Perceived benefits and challenges of working as a teacher - How existing challenges affect teachers attitudes and behaviour at school - Factors that would contribute to doing a better job as a teacher (probe importance of training in relation to other factors) - Satisfaction with acquired skills and skills that can be improved (including gender sensitive pedagogy) - Willingness and ability to use new skills from received training (including gender sensitive pedagogy) - Existing procedures in place to motivate teachers and how useful these are for improving teaching quality - Perceived role and usefulness of teacher appraisals for improved teacher performance (OPRAS and the new TPF)	- KIIs with head teacher and designated INSET teachers (ML and EL only) - KII with community leader - FGDs with teachers - FGD with parents - FGD with SMC - FGDs with pupils - KIIs with key district and ward officials - EQUIP-T programme managers
Output 2: Enhanced school leadership and management skills	- How leadership and management is implemented in the region - How school leaders understand their roles and responsibilities - Areas of satisfaction and frustration experienced by school leaders - How these frustrations affect school leaders attitudes and behaviour at school - How existing challenges can be addressed (probe relevance of, commitment to and effectiveness of EQUIP-T outputs) - Perceived role and impact of appraisals to improve the performance of school leaders (including for improved gender and equity) - Perceived effect of management and leadership training for overcoming the challenges school leaders face (were the training fit for purpose, to what extent has the training been applied, in what ways and with what effect)	- KII with head teacher - KII with community leader - FGDs with teachers - FGD with SMC - KIIs with key district and ward officials
Output 3: Strengthened systems that support the district and regional management of education	- How regional and district management support is implemented in various regions and district - How district and ward officials understand their roles and responsibilities - Challenges district and ward officials face in fulfilling their responsibilities - Effects for schools of existing challenges - Perceived strengths and weaknesses with the district education planning cycle (including planning, budgeting, execution, monitoring) - Perceived frustrations with the management of the capitation grant - Perceived ability to use EMIS data for planning purposes - Factors that would improve the functioning of the district planning cycle - Perceived effect of capacity building activities for improving the planning cycle and releasing capitation grants in full and on time	- KIIs with key district officials - KIIs with Ward Education Coordinators - KII with head teachers - KII with community leader - EQUIP-T programme managers
Output 4: Strengthened community participation and demand for accountability in education	- How community participation and accountability in education is implemented in each region/district - Understandings of the roles and responsibilities of the school committee - What the school committee is good at and where it can improve (including whether represents community views and wishes) - Opinions about the usefulness and impact of training of school committee members - Community/parents views on the access to information about the school and its performance (adequacy, benefits) - Perceived ability of the community/parents to hold school management and teachers to account for the delivery of education - Perceived role and usefulness of the whole school development planning - Views on the relevance and utility of the PTA - Attitudes towards supporting the school (financially, in-kind, other types of support) - Perceived usefulness and transparency of income generating activities - How perceived weaknesses in community participation can be overcome (probe relevance of EQUIP-T activities for improved participation and accountability)	- FGDs with parents - FGDs with pupils - FGD with SMC - FGD with teachers - KII with head teacher - KII with community leader - EQUIP-T Programme managers

4.6   Fieldwork activities

The fieldwork was conducted in two parts in order to balance the need to gather data as soon as possible to maximize the amount of time available for analysis with the need to build on unexpected findings emerging from the quantitative survey.

The first part of the qualitative research took place in June 2014. KIIs were conducted with national, regional, district and ward level stakeholders. The second part of the qualitative research took place in July and August 2014, after the quantitative survey had been completed and preliminary data analysis completed. This allowed the qualitative research team to fully draw on emerging findings from the quantitative survey and explore unexpected or interesting findings from the quantitative baseline survey. This second part of the qualitative research took place at schools and involved school and community level stakeholders.

4.6.1   Addressing rigour through fieldwork

Rigour and the avoidance of bias in the qualitative fieldwork is achieved through extensive training, and the involvement of different individuals in the field teams, so that the teams provide checks on each other. In addition, the teams have kept records of their activities, so that they can be linked to the transcripts and analysis. The following sections outline some of the ways in which rigour was insured, with more detail is provided in Annex J.

4.6.2   The fieldwork teams

The first part of the qualitative fieldwork was conducted by the lead qualitative researcher and a translator for district and ward officials who preferred to carry out the interview in Kiswahili. Interviews conducted in English were transcribed during the interview process with the lead researcher typing the participant responses into a word document, verbatim during the interview process. Interviews conducted in Kiswahili were recorded and the recordings were translated and transcribed post-interview by the interpreter.

For the second part of the qualitative fieldwork, a team of local researchers was trained in Dodoma for a week, commencing on 23 July. The training included a presentation of the key programme objectives and planned interventions, discussions on the TOC of the project, the research questions, and the refining of the research tools, including the participatory tools. The team was also asked to disclose possible biases and discuss how best to ensure that these would not affect the quality of the data. After discussing the research guides and individual questions, and after role playing them in the group, the local researchers 'workshopped' the translation of the guides into Kiswahili. Finally, a one day pilot was conducted as part of the training, after which the instruments were finalised.

One team travelled to all three districts and consisted of the team leader, two moderators, two note takers and one qualitative researcher/interpreter. The team was also accompanied by an OPM staff member who conducted the training and supervised the fieldwork and daily debriefs and also conducting the KIIs.

In each school, four FGDs and two KIIs were conducted. The FGDs were held with:

- Teachers
- School Management Committee members
- Parents of pupils in standard 3

- Pupils in standard 3

The KIIs were conducted with community leaders and head teachers.

4.6.3 Selection of focus group discussion and key informant interview participants

For qualitative research, it is often unrealistic to be too prescriptive over the number of groups and number of participants that must be included in the focus groups. This is because of the range of variables determining participation, particularly with regard to extremely poor and vulnerable people, who are usually less familiar with group discussions and may not be comfortable with participating. In addition, this research is attempting to find explanations for the behaviour of different groups in relation to schooling. This is a very sensitive topic and might further affect participation. Participants were able to come and go from the discussion if needed. The discussions were held where convenient for participants and at times that suited them – i.e. the research team tried to place the needs of participants over the needs of the research team. This was discussed with the head teacher at the initial time of contact by the team leader.

EQUIP-T programme staff were selected by the MA, as the MA is best placed to select the specific programme staff who are knowledgeable on the design, TOC, implementation and progress of the EQUIP-T programme.

REOs and DEOs from each of the sampled regions and districts were approached for interview by a member of the qualitative fieldwork team in advance and interviews were requested. WECs were selected on the basis of being the responsible WEC for each of the schools within the qualitative sample.

Head teachers were informed of the arrival of the team ahead of schedule and were reminded several times of the exact dates for the research. The support of head teachers was enlisted in mobilising key informant interviewees and focus group discussion participants. The head teachers were asked to contact the chairperson of the village committee and if not available a committee member of the village committee for the community leader key informant interview. School management committee members were also invited to participate in the research, as well as teachers and parents of standard III boys and girls. The head teachers were given specific criteria along which to select the parents of standard III pupils, which included sampling both mothers and fathers, parents of both boys and girls and insuring that parents come from diverse socio-economic backgrounds and different religious groups. The participants of the FGD with standard III pupils were the daughters and sons of the parents attending the parent FGD. This approach was adopted in order to ensure informed consent from the parents for the research with standard III pupils (see Annex C on how informed consent was ensured). The participants for the teacher FGD were selected by first selecting all consenting standard I to III math and Kiswahili teachers and then selecting additional teachers in order to compose a group large enough to have a meaningful FGD.

The team had to balance a number of considerations when sampling at the school level. The SMC FGD had to at least contain one teacher and ideally would be made up of an even number of teachers and community members. However, the ability to do so was highly dependent on the size of the teaching body. Several of the schools included in the research have a limited number of teachers at the school which posed challenges for the sampling approach. The chairperson of the village committee could not be included in any of the FGD so that each participant would only be part of one discussion or interview. In addition, as often the case with qualitative research, the sampling was affected by teacher absenteeism and teachers leaving during the research. This resulted in some FGDs not having the desired number of participants.

4.7   Analytical approach

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 be used as a secondary analytical technique, to ensure that the qualitative component is responsive to unexpected information.

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. 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). The set of hypotheses in the qualitative evaluation matrix provided an initial set of themes which were used for confirmatory analysis. These initial themes were used to organise the data into groups (using Excel) that are relevant to these themes and which either confirm or deny the hypotheses. The validity of each piece of data is 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 assess 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. The analysis was conducted by researchers who also conducted fieldwork, thereby helping to ensure that errors of interpretation are minimised.

In addition, new themes were generated from the data through exploratory analysis on the basis of unexpected ideas. These new themes then led to the development of new hypotheses and codes that were applied to the data and tested in similar ways to those set out above under confirmatory analysis. This exploratory analysis is useful, particularly at baseline, to ensure that all relevant themes and hypotheses around EQUIP-T have been identified.

4.7.1   Addressing rigour through the analysis

Rigour in the analysis of the qualitative data comes from four principal sources: first, through the comparison of different data sources, both qualitative and quantitative. Second, different members of the team conducted and discussed the analysis, in order to ensure reliability. Third, the analysis sheet (including de-identified data and the applied data codes) is available for external scrutiny, with confidentiality controls. Finally, the analysis was subjected to peer review and external review.

4.8   Possible limitations of the qualitative component

The main limitations of the qualitative component are presented in Table 9. 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 qualitative data collection and analysis is given below the table.

Table 9 Possible limitations of qualitative component

Possible limitation	Why this is limiting and mitigating factors
Inference beyond the selected research sites is limited.	While the baseline 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 impact evaluation 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 interesting 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.
The qualitative part of the impact evaluation covers all four outputs under outcome one of EQUIP-T and is thus very large in scope which leads to a smaller sample then in a more simplistic design approach	The breadth of the EQUIP-T components to be implemented at multiple levels will necessarily put 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.

The qualitative baseline research explored several sensitive themes, some of which respondents were at first reluctant to discuss. Others such as the differential treatment of girl and boy pupils did not elicit a nuanced response if asked directly and had to be approached in a more subtle way through scenario discussions. The team was trained to anticipate these and trained in how to encourage discussion. For example, participatory tools were available for the team to discuss certain sensitive issues such as gender and teacher and student absenteeism from school.

No clear patterns of reticent responses emerged from the fieldwork debriefs and the data, nor were there specific issues that participants were unwilling to engage with across all groups and cases. However, the cases and respondents differed significantly in their willingness to discuss pupil and teacher absenteeism, especially classroom absenteeism of teachers.

As discussed in the District Management Component section of the Volume I of the main report, REOs and DEOs tended to present resource constraints as the primary challenge to fulfilling their responsibilities. While the researcher probed several times on other challenges, no other explanations were cited.

The most striking observance was the starkly different accounts of the same situation from different groups within cases. For example, parents would frequently report that the relationship between the school and the community was bad, even if head teachers, SC or the community leader had reported that it was good. However, this is perhaps best viewed as a question of difference in perception rather than as a limitation of the actual data.

Bibliography

Abadzi, H. (2006) *Efficient Learning for the Poor: insights from the frontier of cognitive neuroscience.* World Bank, Directions in Development Series.

Adeyemi, T.O. (2010) Principals' leadership styles and teachers' job performance in senior secondary schools in Ondo State, Nigeria *Journal of Education Administration and Policy Studies, 2*(6), 83–91.

Alderson, P. and Morrow, V. (2004) The Ethics of Social Research with Children and Young People. Barkingside: Barnardos.

Attfield, I. and Vu, B.T. (2013) A rising tide of primary school standards—The role of data systems in improving equitable access for all to quality education in Vietnam. *International Journal of Educational Development, 33*(1), 74–87.

Bandur, A. (2012) School-Based Management Developments: Challenges and Impacts. *Journal of Educational Administration, 50*(6), 845–873.

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R. and Khemani, S. (2010) Pitfalls of Participatory Programmes: Evidence from a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy, 2*(1), 1–30.

Berlinski, Galiani and Manacorda, (2008). Giving Children a Better Start: Preschool Attendance and School-Age Profiles. *Journal of Public Economics, 2008, vol.92*, iss.5-6, pp.1416-40.

Birdsall, N., Ibrahim, A. and Gupta, G. (2004) Interim Report on Achieving the Millennium Development Goal of Universal Primary Education United Nations Development Programme.

Blimpo, M. and Evans, D. (2011) School-Based Management and Educational Outcomes: Lessons from a Randomized Field Experiment: The World Bank.

Bruns, B et al (2012). Achieving World-Class Education in Brazil : The Next Agenda. World Bank Report. https://openknowledge.worldbank.org/handle/10986/2383 License: CC BY 3.0 IGO.

Bush, T and Glover, D. (2003) School Leadership: Concepts and Evidence. Summary Report. National College for School Leadership.

Cambridge Education (2014a) Draft EQUIP-T Programme Description, 6 January.

Cambridge Education (2014b) Final EQUIP-T Inception Report, 5 February.

Carvalho, S. and White, H. (1997) 'Combining The Quantitative and Qualitative Approaches to Poverty Measurement and Analysis', Technical Paper 366. Washington DC: World Bank.

Cartwright, N, and Hardie, J., (2012) Evidence Based Policy: A Practical Guide To Doing It Better. Oxford: Oxford University Press.

Casey, K., Glennerster, R. and Miguel, E. (2013) The GoBifo project evaluation report: Assessing the impacts of community-driven development in Sierra Leone (Vol. 3ie Impact Evaluation Report 3).

Cataldo, F., Kielmann, K., and Seeley, J., (2011). Introduction to Qualitative Research Methodology. Research for Development.

CfBT Education Trust (2013) Zimbabwe: School Grants Pilot Baseline Evaluation: www.unicef.org/evaldatabase/files/Zimbabwe_2013-010_Documentation_of_SIG_pilot_Report.pdf

Chang, 2006. National Education Sector Development Plan: A results-based planning handbook. UNESCO IIEP.

Chapman, D., Snyder, C. and Burchfield, S. (1991) Teacher incentives in the Third World (Vol. ERIC Document: ED354621). Washington: Agency for International Development.

Chapman, J. (2005) Recruitment, retention, and development of school principals: UNESCO, International Academy of Education, International Institute for Educational Planning.

Chaudhury, Nazmul, Jeffrey S. Hammer, Michael, Kremer, Karthik Muralidhara, and F. Halsey Rogers. Missing in Action; Teacher and Health Worker Absence in developing countries.

Cogill, J (2008). Primary teachers' interactive whiteboard practice across one year: Change in pedagogy and influencing factors. EdD thesis. King's College University of London.

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.

DataVision (2013) Survey on Flow and Release of School Capitation Grants in Tanzania: Survey Report and Analysis. Dar es Salaam: DataVision International.

Davis, S. H., Leon, R.J. and Fultz, M. (2013) How Principals Learn to Lead: The Comparative Influence of On-the-Job Experiences, Administrator Credential Programmes, and the ISLLC Standards in the Development of Leadership Expertise among Urban Public School Principals. *International Journal of Educational Leadership Preparation, 8*(1).

Day, Sammons, Hopkins, Harris, Leithwood, Gu, Penlington, Mehta and Kington (2007). The Impact of School Leadership on Pupil Outcomes. Interim Report. University of Nottingham.

Dee, T and Wyckoff, J Incentives, Selection, and Teacher Performance: Evidence from IMPACT. NBER Working Paper Series, No. 19529 http://www.nber.org/papers/w19529

Dehejia, R.H. and Wahba, S. (2002). Propensity-Score Matching Methods for Nonexperimental causal studies. *The Review of Economics and Statistics, 84(1): 151–161*

Devarajan, S. (2013) Disrupting Low-level Political Equilibria. from http://blogs.worldbank.org/futuredevelopment/disrupting-low-level-political-equilibria

DFID (2013) Intervention Summary. Education Quality Improvement Programme in Tanzania (EQUIP-T). London, UK: Department for International Development.

Duflo, E. et al (2012) Incentives Work: Getting Teachers to Come to School. American Economic Review 102, 1241–1278.

Duflo, E. et al (2014) School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools. Working Paper.

Dunne, M. and Leach, F. (2005) Gendered School Experiences: the Impact on Retention and Achievement in Botswana and Ghana. London: DFID.

Education Partnerships Incorporated (No date) Coaching Teachers: An Important Principal Role: ERIC document ED538322.

Eller, J.F. (2010) An Evaluation of a Development Programme for New Principals. *The Qualitative Report (ERIC doc: EJ896180), 15*(4), 956–965.

Evans, A. (2014). Does gender sensitisation work? Accessed at https://www.opendemocracy.net/alice-evans/does-gender-sensitisation-work

Evans, L. (1992) 'Teacher Morale: an individual perspective'. *Educational Studies*, 18:2: 161–171.

EQUIP-Tanzania (2014) Draft National School Leadership Competencies Framework

Forum for African Women Educationalist (FAWE) (1998). Experiences in Creating a Conducive Environment for Girls in School.

Furman, G. (ed.) (2012) *School as community: From promise to practice*: SUNY Press.

Galab, S. et al (2013) Community-based accountability for school improvement.

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.

Gertler, P.J., Martinez, S., Premand, P., Rawlings, L. and Vermeersch, C. (2011) Impact Evaluation in Practice. Washington, DC: World Bank.

Gertler, P et al (2012) Parental Empowerment in Mexico: Randomized Experiment of the "Apoyos a La Gestion Escolar (Age)" Program in Rural Primary Schools in Mexico. Society for Research on Educational Effectiveness.

Getzel, J. W., & Guba, E. G. (1957). Social behaviour and the administrative process. *The School Review, 65*, 423-441.

Glewwe, P., Maïga, E.W.H. (2011) The impacts of school management reforms in Madagascar: do the impacts vary by teacher type? Journal of Development Effectiveness 3, 435–469. doi:10.1080/19439342.2011.604729

Global Partnership for Education (2013) All Children Learning Report.

Government of Tanzania (2007) Mtaala wa elima ya msingi. Dar es Salaam.

Gowlett (2011) Injurious assumptions: Butler, subjectification and gen(d)erational poverty. In Outhred, R. Reconceptualising Equity in Higher Education. *International Journal of Inclusive Education. Volume 16. No 9.*

Guest, G., MacQueen, K., Namey, E., (2010). Applied Thematic Analysis. Sage Publications.

Guryan, J. (2001) Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts *NBER Working Paper No. 8269*: National Bureau of Economic Research.

Haegeland, T., Raaum, O. and Salvanes, K.G. (2012) Pennies from Heaven? Using Exogenous Tax Variation to Identify Effects of School Resources on Pupil Achievement. *Economics of Education Review, 31*(5), 601–614.

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.

Hare, H. (2007) ICT Education in Tanzania. *Survey of ICT and Education in Africa.* from www.infodev.org/infodev-files/resource/InfodevDocuments_432.pdf

Hattie, J. (2003) Teachers Make a Difference. What is the research? Distinguishing expert teachers from novice and experienced teachers. University of Auckland. Australian Council for Educational Research.

Heinrich, C, Maffioli, A. and Vázquez, G. (2010). A Primer for Applying Propensity-Score Matching. *Impact-Evaluation Guidelines Technical Notes No. IDB-TN-161*. Washington, DC: Inter-American Development Bank.

Heyneman, S.P. (2013) The International Efficiency of American Education: The Bad and the Not-So-Bad News. *Current Issues in Comparative Education, 16*(1), 74–89.

Holden, K.L. (2013) Buy the Book? Evidence on the Effect of Textbook Funding on Student Achievement. *Working Papers.* from http://pages.uoregon.edu/kholden/uploads/1/3/8/7/13879025/bythebook_holden.pdf

Hopkins, J., Harris, A., and Jackson, D. (1997). Understanding the School's Capacity for Development: Growth states and strategies. *School Leadership and Management, 17*(3), 401-412.

Jull, S., Swaffield, S. and MacBeath, V. (2014) Changing perceptions is one thing…: barriers to transforming leadership and learning in Ghanaian basic schools. *School Leadership & Management, 34*(1).

Kane, E, (1996). Gender, Culture and Learning. Advancing Basic Education and Literacy, Phase 2. Education Development Centre.

Khandker, S.R., Koolwal, G.B. and Samad, H.A. (2010) *Handbook on Impact Evaluation. Quantitative Methods in Practices.* Washington, DC: World Bank.

Kiggundu, E., Moorosi, P. (2012) Networking for school leadership in South Africa: perceptions and realities. School Leadership & Management 32, 215–232. doi:10.1080/13632434.2012.688738

Koerner T. F. (1990). Developing staff morale. *The Practitioner, 16*, 4.

Korach, S. and Agans, L.J. (2011) From Ground to Distance: The Impact of Advanced Technologies on an Innovative School Leadership Programme. *Journal of Research on Leadership Education, 6*(5), 216–233.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. (2009). "Incentives to Learn." *The Review of Economics and Statistics 91*(3): 537-56.

Krippendorff, K., (2004). Content Analysis: An Introduction to Its Methodology. Sage Publications.

Kunje, D., Selemani-Meke, E. and Ogawa, K. (2009) An Investigation of the Relationship between School and Pupil Characteristics and Achievement at the Basic Education Level in Malawi. *Journal of International Cooperation in Education, 12*(1), 33–49.

Leach, J (2002). The curriculum knowledge of teachers: a review of the potential of large-scale electronic conference environments for professional development. *Curriculum Journal, 13*(1), pp. 87–120.

Lee, V.E. and Zuze, T.L. (2011) School Resources and Academic Performance in Sub-Saharan Africa. *Comparative Education Review, 55*(3), 369–397.

Leithwood. (2010) Reasons for school failure. Chapter 1 in *Leading school turnaround.*

Lewin, K. (2010) Looking back to see the future: Four decades of development in educational planning. In M. Bray and N. V. Varghese (eds.), *Directions in educational planning; international experiences and perspectives.* Paris: UNESCO IIEP.

Lipham, J. M., Rankin, R. E., & Hoeh, J. A. (1985). The principalship: Concepts, competencies, and cases. New York: Longman.

Lumadi, M., L. (2010). The Need For Training Gender-Sensitive Teachers: Addressing Education Challenges for Gender Sensitive National Development. Contemporary Issues in Education Research. March 2010 Volume 3, Number 3.

Manara, K. and Mwombela, S. (no date) The Governance of the Capitation Grant in Primary Education in Tanzania: Why Civic Engagement and School Autonomy Matter. In REPOA (ed.). Dar es Salaam: REPOA.

Mandebvu, O.S. and Chitekuteku, S.R. (2000) An Evaluation of the Managerial Skills Training Programme for Primary School Heads in Zimbabwe. Harare Regional Office: UNICEF.

Mannathoko, C. E. (1994) Democracy in the management of teacher education in Botswana. British Journal of Sociology of Education, Volume 15(4), pp.481-495.

Martinex, S, Naudeau, S and Pereira V. (2012). The Promise of Preschool in Africa: A Randomized Impact Evaluation of Early Childhood Development in Rural Mozambique. 3ie.

Mathaba, R.S.R. (Undated) *External whole school evaluation of underperforming secondary schools in Mpumalanga Province.* (Doctor of Technology), Durban University of Technology, South Africa. Retrieved from http://ir.dut.ac.za/bitstream/handle/10321/1073/MATHABA_2014.pdf?sequence=1

Mays, N., and Pope, C. (1995). Rigour and Qualitative Research, British Medical Journal 311(6997): 109-112.

McEnrue, M.P. (1984) Perceived competence as a moderator of the relationship between role clarity and job performance: A test of two hypotheses. *Organizational Behavior and Human Performance, 34*(3), 379–386.

McKenzie, D. and Woodruff, C. (2012) What are we learning from business training and entrepreneurship evaluations around the developing world? *Policy Research Working Paper 6202*: The World Bank Development Research Group.

Mendel, P. C. (1987). An investigation of factors that influence teacher morale and satisfaction with work conditions. Unpublished Dissertation, University of Oregon, Eugene.

Mestry, R. and Schmidt, M. (2010) Portfolio Assessment as a Tool for Promoting Professional Development of School Principals: A South African Perspective. *Education and Urban Society, 42*(3), 352–373.

Militello, M., Fusarelli, B., Alsbury, T. and Warren, T.P. (2013) How Professional Standards Guide Practice for School Principals. *International Journal of Educational Management, 27*(1), 74–90.

Ministry of Labour. (2003). United Republic of Tanzania: National Ageing Policy. Ministry of Labour, Youth Development and Sports.

Minnich, E. K. (1990). Transforming knowledge. Philadelphia, PA: Temple University Press.

MoEVT (2009a) Basic standards for pre- and primary education in Tanzania. Dar es Salaam: Ministry of Education and Vocational Training. www.ed-dpg.or.tz/pdf/PE/Basic%20Standards%20for%20Pre%20and%20Primary%20Education%20in%20TZ_2009.pdf

MoEVT (2009b) *Handbook for School Inspectors Third Edition 2009*. Dar es Salaam: Ministry of Education and Vocational Training. www.ed-dpg.or.tz/pdf/PE/Inspector's_Handbook_Final%20March%202010.pdf

MoEVT (2009c) *Guidelines for School Supervision.* September, 2009. Dar es Salaam: Ministry of Education and Vocational Training. www.ed-dpg.or.tz/pdf/PE/School%20Supervision_guidelines_Final%20March%202010.pdf

Molander, C. and Winterton, J. (1994) *Managing Human Resources.* London: Routledge.

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

Muzale and Rugemalira (2008). Researching and Documenting the Languages of Tanzania. Language Documentation and Conservation. Vol. 2 pp 68-108.

Ng, S.W. (2013) Equipping Aspiring Principals for the Principalship in Hong Kong. *Educational Management Administration & Leadership, 41*(3), 272–288.

OECD (2008) Improving School Leadership Volume 1: Policy and Practice. www.oecd.org/education/school/41200009.pdf

Ombui, S.C. (2009) Challenges facing principals in Integrating Information Communication Technology (ICT) for efficient management in secondary schools In Nyamira County, Kenya. from http://ir-library.ku.ac.ke/handle/123456789/7267

OPM (2014a) 'Impact Evaluation of the Education Quality Improvement Programme in Tanzania (EQUIP-T): Inception Report and Baseline Design'. Oxford, UK: Oxford Policy Management.

OPM (2014b) 'EQUIP-T Impact Evaluation. Preliminary Findings from the Baseline Quantitative Survey for Programme Treatment Districts'. Draft Report. Oxford, UK: Oxford Policy Management.

Osei, R., Owusu, G., Asem, F. and Afutu-Kotey, R. (2009) Effects of capitation grant on education outcomes in Ghana: Institute of Statistical Social and Economic Research.

Oulai, D., Lugaz, C., Minas, A. and Teklehaimanot, H. (2011) Analysis of capacity development in educational planning and management in Ethiopia. Paris: International Institute for Educational Planning.

Palomer, C.G. and Paredes, R.D. (2010) Reducing the Educational Gap: Good Results in Vulnerable Groups. *Journal of Development Studies, 46*(March), 535–555.

Pandey, P. et al (2009) Community participation in public schools: impact of information campaigns in three Indian states. *Education Economics 17*, 355–375. doi:10.1080/09645290903157484

Pandey, P., Goyal, S. and Sundararaman, V. (2011) Does Information Improve School Accountability? Results of a Large Randomized Trial (Vol. Report No. 49): International Bank for Reconstruction and Development/ and The World Bank – South Asia: Human Development Unit.

Pole de Dakar (2012) Tanzania Education Sector Analysis. Beyond Primary Education, the Quest for Balanced and Efficient Policy Choices for Human Development and Economic Growth. Geneva: United Nations Educational Scientific and Cultural Organization.

Ponera, G.E., Mhonyiwa, J.E. and Mrutu, A.S. (2011) Quality of Primary School Inputs in Tanzania Mainland. Policy Brief Number 2 (October 2011): www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/policy-brief/tan_school_inputs_15oct2011_final.pdf [09 September 2014

Pradhan, M. et al. (2011) Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia *Policy Research working paper; no. WPS 5795*. Washington, DC: The World Bank.

Prendergast, C. Intrinsic motivation and incentives. *Am Econ Rev* 2008; 98: 201-5 doi: 10.1257/aer.98.2.201.

Pritchett, L. and Beatty, A. (2012) 'The negative consequences of overambitious curricula in developing countries' *Faculty Research Working Paper Series RWP 12-035*. Harvard Kennedy School.

Ramanujam, V. and Venkatraman, N. (1987) Planning System Characteristics and Planning Effectiveness. From http://dspace.mit.edu/bitstream/handle/1721.1/48833/planningsystemsc00rama.pdf?sequence=1&origin=publicationDetail

Richards, J.C. and Theodore, S (2001) Approaches and Methods in Language Teaching (2nd Ed.). Cambridge, New York: Cambridge University Press.

Robinson, V.M. (2007) School Leadership and Student Outcomes: Identifying what works and why *ACEL Monograph series* (Vol. 41).

Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. *Education Administration Quarterly, 44*(5), 635-674.

Romano, Babchishin, Pagani and Kohen (2010). School readiness and later achievement: replication dn extension using a nationwide Canadian survey. *Development Psychology. September 46* (5) 995-1007.

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.

Sampa, P.C. et al (2009) Improving quality in Zambia's community schools: Sustainable strategies or quick fixes? Working Paper.

Scheerens, J. (2004) The quality of education at the beginning of the 21st Century: UNESCO.

Schenk, K. and Williamson, J. (2005) Ethical Approaches to Gathering Information from Children and Adolescents in International Settings: Guidelines and Resources. Washington, DC: Population Council.

Schreiner, M. (2013) 'A Simple Poverty Scorecard for Tanzania.' Kansas City: Microfinance Risk Management, L.L.C. www.microfinance.com/#Tanzania

Sumra, S, and Katabaro, J.K. (2014) Declining Quality of Education: Suggestions for Arresting and Reversing the Trend. Special THDR Issues. ESRF Discussion Paper 63. The Economic and Social Research Foundation, Dar es Salaam.

Spaull, N and Viljoen, J. (2013) Starting Behind and Staying behind: the case of insurmountable learning deficits in South Africa. Draft.

Stephens, C. (2013) EQUIP-T Impact Evaluation: Key Literature Analysis. Draft, December.

Swallow, J. et al (2009) USAID/Zambia Education Program Evaluation. USAID.

Tembo, F. (2013) Rethinking social accountability in Africa: lessons from the Mwananchi Programme: UKAid.

UNESCO (2010) Assessing Education Data Quality in the Southern African Development Community (SADC): A Synthesis of Seven Country Assessments: UNESCO.

UNESCO (2011). Financing education in sub-Saharan Africa. Meeting the challenges of expansion equity and quality.

UNESCO (2012a) EFA Global Monitoring Report 2011. The hidden crisis: Armed conflict and education. Geneva: United Nations Educational Scientific and Cultural Organization.

UNESCO 2012 (2012b). World Atlas of Gender Equality in Education.

UNESCO (2014) EFA Global Monitoring Report teaching and learning: Achieving quality for all: UNESCO.

UNESCO-IIEP (2011). Gender Equality in Education. Looking beyond Parity. An IIEP Evidence-Based Policy Forum. Paris, France. 3-4 October 2011.

UNESCO IIEP (no date). Educational Planning: Approaches, challenges and international frameworks. Distance Education Programme on Education Sector Planning. Module I. Accessed at http://www.moe.gov.jo/Files/(22-12-2013)(1-01-43%20PM).pdf

UNESCO PROAP, (2000). Increasing the Number of Women Teachers in Rural School. A Synthesis of Country Case Studies. South Asia. UNESCO Principal Regional Office for Asia and the Pacific and Asia-Pacific Programme of Education for All.

UNICEF (2002) Children Participating in Research, Monitoring and Evaluation (M&E): Ethics and Your Responsibilities As A Manager. Evaluation Technical Notes No 1. Geneva: UNICEF Evaluation Office.

URT (2010) 'Public Expenditure Tracking Survey of Primary and Secondary Education in Mainland Tanzania.' Final report, February 2010. Dar es Salaam: Ministry of Education and Vocational Training.

UWEZO (2011) Are Our Children Learning? Annual Learning Assessment 2011. Dar es Salaam: UWEZO Tanzania.

UWEZO (2013) Are Our Children Learning? Annual Learning Assessment Report 2012. Dar es Salaam: UWEZO Tanzania.

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.

Wamakote, L., Ang'ondi, E. and Onguko, B. (2010) East African Government Investments in Information and Communication Technologies for Education: Matching Policy to Practice. *Itupale Online Journal of African Studies, 2*, 25–38.

Washington, R., & Watson, H. F. (1976). Positive teacher morale: The principal's responsibility. NASSP Bulletin 60, 4-6.

Waters, T., Marzno, R. and McNulty, B. (2003) Balanced Leadership: What 30 years of research tells us about the effect of leadership on student achievement: McRel.

Westhorp, G., Walker, B., Rogers, P., Overbeeke, N., Ball, D., and Brice, G. (2014). Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: A realist review. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

What Works Clearinghouse (2012) WWC review of the report: Teacher incentive pay and educational outcomes: Evidence from the New York City Bonus Program. U.S. Department of Education, Institute of Education Sciences, Retrieved from http://whatworks.ed.gov.

White, H. (2009) Theory-Based Impact Evaluation: Principles and Practice, 3ie Working Paper 3. *International Initiative on Impact Evaluation.*

Woolston, W. 2008. Do "Great Expectations" Matter? The relationship between teacher expectations and student academic success. The Institute for Research on Education Policy & Practice. Stanford University.

World Bank (2010) Does linking teacher pay to student performance improve results? (Report No. 57782).

World Bank (2012) 'Service Delivery Indicators. Tanzania.' Washington, DC: World Bank.

Wright, B and Stone, M. (1979) *Best Test Design.* Chicago: MESA Press.

Zachariou, A., Kadji-Beltran, C. and Manoli, C.C. (2013) School principals' professional development in the framework of sustainable schools in Cyprus: a matter of refocusing. *Professional Development in Education, 39*(5).

Zuga, K. (1999). Addressing Women's Ways of Knowing to Improve the Technology Education Environment for All Students. *Journal of Technology Education. Volume 10, Number 2*.

Annex A   Impact evaluation 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 will only be part of 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
	Mara	Bunda DC
		Butiama DC
		Musuma DC
		Musuma MC
		Rorya DC

Annex B   Stakeholder consultations

During the inception phase the impact evaluation team spent time in Tanzania to meet with stakeholders and collect data and information for the design process and to prepare for implementation. An overview of these consultations and activities is provided in Annex table 2 (for further details see OPM 2014a).

Annex table 2 Stakeholder consultations and missions

Date	Purpose
September 2013	• Discuss the impact evaluation design with DFID • Meet with national education stakeholders to introduce the impact evaluation of EQUIP-T and collect information to inform the design process • Meet with the MA to receive information on the EQUIP-T programme design and activities • Collect information on pupil learning assessment conducted in Tanzania previously • Meetings to understand data availability for construction of the district and school sampling frames • District and school visits to begin to understand education constraints at these levels and to introduce the impact evaluation
November/December 2013	• Receive the EQUIP-T programme design information from the MA • Discuss the impact evaluation design and timetable with DFID • Meet with MoEVT to discuss design developments and collect further information • School visits to inform the quantitative instrument design • School visits to guide the development of the qualitative research design • Meetings with ADEM, MoEVT and University of Dar es Salaam (UDSM) staff to gather information, especially on gender issues in education, to inform the qualitative research design • Identification of potential national researchers for the qualitative fieldwork • Meetings to further inform the development of the pupil learning assessment • Collection of EMIS and 2012 Population Census data
December 2013-January 2014	• Development of pupil and teacher learning assessments by a national team working closely with the impact evaluation team • Pre-test 1 of paper quantitative instruments
January 2014	• Qualitative follow up scoping mission • Interviews with national qualitative researchers for the qualitative fieldwork
January 31, 2014	• First Reference Group for the Impact Evaluation meeting to present and receive technical feedback on the summary draft impact evaluation design and the survey instruments
February 2014	• Pre-test 2 using CAPI survey instruments
November 2013-March 2014	• Identification and recruitment of analysts, field supervisors and enumerators for the quantitative baseline survey by the OPM Tanzania Office • Development of fieldwork model for quantitative baseline survey by the OPM Tanzania Office • Pre-test 3 using CAPI survey instruments by the impact evaluation team • Supervisor and enumerator training and pilot for quantitative baseline survey by the impact evaluation team
May 2014	• Collect EQUIP-T MIS data on programme costs for the impact evaluation fiscal affordability study • Meetings with stakeholders including DFID, MoEVT, PMORALG and the MA to discuss preliminary and partial findings from the quantitative baseline survey • Scoping to finalise revised qualitative research design
August 2014	• Visit to supervise final stages of data checking and cleaning process
June and July-August 2014	• Training, piloting and qualitative fieldwork in two stages
November 2014	• Second Reference Group meeting • Presentation of findings at EQUIP-T Annual Review, and to DFID office
February 2015 (preliminary)	• Dissemination of impact evaluation findings at Annual Education Sector Review meeting

Annex C   Permits, consent, confidentiality and datasets

Conducting fieldwork requires high ethical standards to ensure that expectations are not raised, confidentiality is maintained and respondents are never forced to participate or encouraged to speak about subjects that may be traumatising (especially for children).

The key areas for ethical consideration for research are: (i) informed consent; (ii) harms and benefits; (iii) payment and compensation; and (iv) privacy and confidentiality. These areas are addressed below.

The quantitative fieldwork was carried out by field teams made up of national enumerators and field supervisors supported by OPM staff from the OPM Tanzania office. The interviews with head teachers, teachers and pupil testing were conducted in Kiswahili. The OPM Tanzania office team after discussions with the MOEVT arranged for delivery of letters of permission to visit schools to the DEOs in all the selected survey districts. Sending the permit letters was not considered sufficient to ensure the DEOs had read and agreed to the school visits. Therefore, follow-up phone calls were made to confirm that the DEOs had received the letters seeking permission to visit schools from the MoEVT and understood the research purpose and allowed the field teams to visit schools in their district. Arriving at schools, the team supervisors started by introducing themselves and their teams to the head teacher, explaining the purpose of the visit and time that would be required to complete the survey. The enumerators introduced the study and interviews/texts to the head teacher and to all the respondents (pupils and teachers. All respondents were given the option to refuse to participate in the study. If a respondent was reluctant and or more explanations were requested, the enumerators had been trained to be as exhaustive as possible in explaining the study and its purpose. No head teacher or pupil declined to participate in the survey, and only in a few rare cases did a teacher decline to complete the TDNA citing time constraints.

The qualitative fieldwork was carried out by a team made up of national researchers and OPM staff. KIIs and FGDs were conducted in Kiswahili. The field teams made all efforts possible to keep disruptions of the school day to a minimum by ensuring that head teachers were informed as early as possible of the dates of the school visits and which types of FGDs and KIIs would take place. The FGDs and KIIs were recorded after the informed consent of participants was granted. The sequencing of KIIs and FGDs was also - as far as possible - organised in cooperation with community and school members in order to minimise disruption to school and village life and ensure smooth running of the research. KIIs frequently took place outside the school in order to minimise disruption within teaching spaces.

Informed consent was sought from all participants at the national, regional, district, ward, community and school level for the qualitative research. The aims of the research and their ability to withdraw consent at any point during the interviews or discussions were explained to participants. At the school and community levels, registries were completed with participants where participants either signed or made a stamp to indicate their willingness to participate in the research. In order to ensure that participants were comfortable with the procedure researchers would read out the explanation and ask participants whether the information provided was clear. Participants were invited to either end or temporally interrupt the interview or discussion if additional questions or concerns arose.

The fieldwork included FGDs with children. The children participating in the research were boys and girls from standard three. There is some debate in the development community and within Tanzania regarding who is in a position to provide consent for research conducted with participants who are young children. For ethical reasons, the team decided to gain both the consent of parents and children. As a first step, parents of standard three boys and girls were asked to provide written

consent that they were willing to allow their children to participate in the research. If permission was granted, the children were asked to also provide written consent of their willingness to participate in the research. At both stages the nature of the research was explained and it was made clear that both parents and children were under no obligation to participate. As there were some parents and pupils who did not wish to participate and felt comfortable communicating this decision to the field team this suggests that the approach taken to obtain informed consent worked.

No incentives were given to respondents for participation in the study. Each school that was part of the quantitative baseline survey received a gift (inflatable world map) and participants in the FGDs and KIIs for the second part of the qualitative fieldwork (schools and communities) received refreshments and the schools received crayons, pencils and chalk.

All personal data collected as part of this survey are only available to authorised individuals for analytical purposes and are handled using data protection best practices. Each respondent has been assigned a unique identifier that is used to analyse the data. All cleaned and documented datasets, anonymised by removing personal information that could be used to identify respondents, related to the baseline study will be made public (subject to DFID approval) to enable national researchers, research students and other education stakeholders to access and use the IE data to conduct additional analysis and research. All data have been backed up and are stored in an 'OPM Stats archive'.

During the preparatory process for this study, the ethical considerations by the EQUIP-T IE Reference Group were based on enquires into standard ethical questions related to the conduct of research with children and vulnerable groups, and were included in the researcher manual and covered in researcher training. These are summarised in the table below.

Annex table 3 Ethical and child protection considerations

Ethical and child protection considerations

How are participants being selected? Is there any deliberate exclusion on the basis of, for example, access or stigma? Have cultural and community norms been understood and considered in the selection process? People should be offered the opportunity to participate and, ideally, be invited to volunteer, rather than be asked to avoid any pressure. People have the right not to participate and to pull out at any stage.

Ensuring that permission is sought for the focus groups to go ahead, through consultation with the local community and school.

Setting and communicating clear parameters for the focus group – this means clearly stating the purpose, the limits and what the follow up will entail. It also means ensuring that demands on participants' time are not excessive and that they are aware of their right to not participate or withdraw at any time.

Recognising that participants are possibly vulnerable and that the exercise is carried out with full respect – power differentials will exist between community members, within schools and researchers: for example people can easily be treated as inferior and such power dynamics need to be understood and purposefully mitigated in planning and implementation by researchers and facilitators.

Ensuring the safety and protection of participants – this means ensuring the environment is physically safe, that there are at least two facilitators present at all times. Facilitators should also be supervised.

Ensuring that people understand what is happening at all time. Is appropriate language being used (language, dialect, community terminology, etc)? This needs to be carefully planned.

Ensuring the right to privacy – this includes ensuring anonymity and confidentiality, in record keeping and report writing and making sure participants understand that what they do and say in the group session will remain anonymous.

Ensuring that plans are in place for dealing with unexpected or adverse consequences, such as an urgent physical protection issue. Provision of immediate support needs to be planned for in advance.

The following ethical principles are particularly relevant when working with children:

- **Is participation of children genuinely necessary**: is participation strictly necessary to meet the needs of the research enquiry?
- **Recognising that child participants are vulnerable**: the exercise should be carried out with full respect – children for example can easily be treated as inferior and such power dynamics need to be understood and purposefully mitigated in planning and implementation by researchers, enumerators and facilitators.
- **Ensuring that the consent of appropriate adults** (parents, head teacher) is sought for a child's participation: as well as ensuring that the child gives their own informed consent to take part (is happy to do and is not coerced in any way). Children's participation in research must be fully informed and children must also be free not to participate.
- **Ensuring confidentiality**, except in circumstances where a child discloses information about a serious child protection concern: in which case the researchers may need to act (reporting, immediate protection).
- Ensuring safety of child participant: If this is in any doubt, at any stage, the activity should be halted.

Annex D   Dissemination

The IE aims to disseminate the findings to key audiences including DFID, the EQUIP-T MA, the Government, sub-national level education stakeholders comprising REOs, DEOs, WECs, schools and communities, and academia.

As part of the IE dissemination plan, different products will be developed for different audiences. These include this baseline report which is comprehensive and technical in nature; thematic reports on specific topics that will be shorter and policy oriented; policy leaflets that outline the key evaluation findings for a non-technical audience; and presentations to be used by the IE team for dissemination and learning purposes. Reports, leaflets and presentations will be posted on the OPM website (subject to DFID approval) to grant easy access for education stakeholders in Tanzania and elsewhere.

The OPM IE team will seek to use multiple channels to stimulate a wider, evidence-based discussion of the key issues through workshops, one-on-one demand-driven briefings with MoEVT, PMO-RALG and other key national stakeholders, presentations at education events and the use of existing government dissemination channels. In particular, the OPM impact evaluation team will present the baseline results at the EQUIP-T programme review in Kigoma and hold the second Reference Group meeting in November 2014. The team will also present at the Joint Education Sector Review meetings in 2015, 2016/17 and 2018/19. Additional channels for results dissemination include the MOEVT bi-weekly media slots and national media.

To reach WECs, schools and communities the IE team will discuss with the EQUIP-T MA and UWEZO, which both have sub-national networks, the possibility of using their communication channels to disseminate key findings for each programme region in a non-technical and concise format.

Annex E   Quantitative data collection, cleaning and analysis

The baseline survey was conducted by OPM's Tanzania office working closely with the OPM IE team that designed the IE and conducted the analysis, throughout the baseline.

E.1   Permits

In order to visit schools in Tanzania a permit from the MOEVT is required. The OPM Tanzania office obtained permission letters for all the 25 IE districts from MOEVT and sent these to each of the DEOs and followed up with phone calls to confirm receipt and that the purpose of the baseline survey was well understood and accepted.

E.2   Recruitment of enumerators and field teams

The recruitment process was conducted to ensure that a quality pool of enumerators and supervisors would be selected. To get candidates with a suitable background priority was given to candidates with prior experience in conducting education surveys and background in teaching. All potential candidates received an overall score based on: survey experience (with a particular focus on experience within education), education and appropriate attitude. Good understanding of English language was an essential requirement. Further, candidates were selected into groups of potential supervisors, educators, and enumerators given emphasis to the key skills needed for each role.

51 people were trained to ensure selection of the best candidates and to have a reserve to select from if needed during the fieldwork. The final field teams were selected at the end of the training based on: active participation during the training, ability to follow field work procedures, and administer interviews during the training practice days, ability to use CAPI and availability throughout the fieldwork: 45 enumerators (including educators) and team supervisors were selected. Each baseline field team consisted of:

- One team supervisor;
- Two enumerators (of which one was an educator responsible for conducting the lesson observations); and
- One driver.

E.3   Training and pilot

Training of enumerators for the fieldwork was conducted in Morogoro region March 13-22, 2014 led by a team comprising OPM IE team members and OPM the Tanzania office survey team. The training was split into three parts:

- Training of potential team supervisors and educators on how to conduct the head teacher interview (conducted by supervisors) and the lesson observation (conducted by educators);
- Training of all potential enumerators in using the CAPI software and tablets including in the field data validation, the data management procedure including how to send the data to the OPM Tanzania office, data, overall fieldwork procedures and instruments (excluding the head teacher interview and lesson observation): and
- A two day pilot in Singida region.

The training had two modalities: i) 'classroom-based' interspersed with ii) field practice. During the class room sessions each question of the eight instruments was presented and procedures explained. Different training tools were used such as videos from lesson observations, pictures taken at pre-test schools and role play. Feedback from the enumerators were collected daily by OPM staff and addressed to further improve instrument translation, the enumerator etc. The instruments covered during each classroom-based session were tested in the field the following day(s) to be able to practice what had been learnt in the classroom. OPM staff supervised the teams during the field practice and debriefs were held after.

The pilot was conducted immediately after the last day of training in Singida region. The objective was to ensure that i) supervisors learnt how to organise themselves and the team members in the field and ii) that supervisors and enumerators experienced a full two days in a school practicing how to collect all the required data within the allocated timeframe.

During the first two days of fieldwork six OPM survey team members supervised the teams in the field and held daily debrief sessions to discuss challenges faced and ensure that supervisors were able to send the collected data back to the OPM Tanzania office.

E.4   Enumerator manual

To provide fieldworkers with a reference tool as well as a support during data collection, a comprehensive enumerator manual for the fieldwork implementation was developed. The manual covers codes of conduct, the use of CAPI, data validation and transfer procedure, instructions on sampling within schools amongst other things as well as instrument description and instructions how to administer each instrument.

E.5   Fieldwork planning and implementation

The fieldwork covered 25 districts in 12 regions in Tanzania: Arusha, Dodoma, Kigoma, Mwanza, Pwani, Rukwa, Ruvuma, Shinyanga, Simiyu, Singida, Tabora and Tanga. The sample consisted of 200 primary schools (see Table 1 for an overview).

The field work model and execution required detailed planning. All schools had to be visited within a short timeframe from end of March to mid-May. The sampled schools had mid-term breaks on different dates depending on the region as well as mid-term examinations leading up to the mid-term break. The plan had to be flexible to allow for revision whenever an unscheduled event like road destruction, floods, or out of school sports events and other activities occurred.

The fieldwork was conducted during a seven week window March 27-May 13, 2014. One field team (see section E.3 for team composition) was able to complete data collection for one school in two days on average.

Due to insurmountable logistical challenges or activities in schools that prevented observation of regular school activities such as vaccination campaigns, five schools (one in Arusha, one in Singida, one in Shinyanga and two in Tanga,), were replaced using the reserve list prepared in advance.

E.6   Quality assurance during fieldwork

The OPM Tanzania office team put several QA mechanisms in place to ensure that data with high quality was collected. The mechanisms included:

- Supervision of all teams during the roll out of fieldwork by members from the OPM Tanzania office survey team;

- Supervision of all teams by an external survey expert, the fieldwork QA supervisor, and assistant survey manager during the second week of fieldwork, and fieldwork supervision by the fieldwork QA supervisor throughout the data collection phase;

- Two additional training sessions were held over weekends to teams that were identified as needing extra support;

- Additional QA visits as needed from the OPM Tanzania office survey manager and assistant survey manager to support relatively weaker teams; and

- Continuous support and feedback from the OPM Tanzania offices survey team via phone and email.

For the duration of the fieldwork one OPM staff member worked full time, travelling between the twelve regions, to supervise the field teams. In addition, the QA fieldwork supervisor visited most teams at least once, to provide support and monitor the field teams' work.

E.7   Data validation, cleaning and analysis

E.7.1   Data validation during the fieldwork

A clear advantage of using CAPI for the survey work was that the data was received almost immediately from the field by the OPM Tanzania office survey team in Dar es Salaam so that identified fieldwork errors could be corrected while teams were still in the field. Every two days the 15 field teams sent a maximum of 50 data files each back to the OPM Tanzania office and the assistant data manager validated and checked the quality of the data received. This included: i) that the correct number of data files had been received; ii) opening each file in the custom-designed CAPI software to validate the data according to the built-in flags and checks, and iii) communicate to team supervisors how to correct any detected data issues.

E.7.2   Data checking and cleaning 1

All sampled schools, head teachers, teachers and pupils were uniquely identified by ID codes assigned either before the fieldwork (region, district and school IDs), or at the time of the school visit using automated tables in CAPI (teacher, lesson observation and pupil IDs). The first set of checking activities included checking: of all IDs, for missing observations, and for missing responses where none should be missing. This resulted in a raw dataset sent to the analysis team.

E.7.3   Data checking and cleaning 2 and analysis

A comprehensive data analysis system was created including a logical folder structure, the development of a detailed data documentation guide and template syntax files, to ensure all analysts used the same file and variable naming conventions, variable definitions and disaggregation variables, and weighted all estimates appropriately.

The raw datasets were subject to a second set of detailed checking and cleaning by the analysis team. The final, clean datasets including the generated IE indicators were constructed and analysed in Stata using standardised methods and clear documentation.

The entire analysis was implemented taking into account the sampling structure of the survey. In Stata, the 'svyset' and 'svy' commands were used to ensure that all estimates incorporated the correct survey set-up and weights (see section 3.2.4). In order to test whether differences

observed between groups are statistically significant, a Wald test adjusted for sampling weights was carried out for each indicator.

Annex F   Definitions of key quantitative indicators

Annex table 4 Definition of key indicators

Indicator name	Indicator definition	Respondent/observation	Notes
Chapter 2: Pupil learning			
Pupil learning Kiswahili			
Pupil ability score (logits)	Estimate of pupil ability in Kiswahili on a linear scale.	Standard 3 pupils	Estimates of pupil ability and item difficulty are estimated using Rasch analysis (a form of 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.
Pupils in performance band x (%)	Number of pupils with ability scores which fall in the band boundaries / total number of pupils, expressed as a percentage. Note: the 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.	Standard 3 pupils	
Number of words read per minute	Number of words correctly read by pupil / time taken to read the words in minutes (note: pupils were given one minute to complete each oral test; a few pupils finished in less than one minute)	Standard 3 pupils	These indicators are constructed using raw-score data.

Non-response' was treated as incorrect in the results. Most non-response happened because of instructions in the test to skip questions if a pupil got a fixed number of prior questions incorrect. 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. |
Test score (%)	Number of questions answered correctly / total number of questions, expressed as a percentage, for each pupil	Standard 3 pupils	
Percentage of pupils who scored more than x%	Number of pupils who scored more than x% of questions correct / total number of pupils, expressed as a percentage	Standard 3 pupils	
Pupil learning mathematics			
Pupil ability score (logits)	Estimate of pupil ability in mathematics on a linear scale.	Standard 3 pupils	Estimates of pupil ability and item difficulty are estimated using Rasch analysis (a form of 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.
Pupils in performance band x (%)	Number of pupils with ability scores which fall in the band boundaries / total number of pupils, expressed as a percentage. Note: the 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.	Standard 3 pupils	

Test score (%)	Number of questions answered correctly / total number of questions, expressed as a percentage, for each pupil	Standard 3 pupils	These indicators are constructed using raw-score data.
Percentage of pupils who scored more than x%	Number of pupils who scored more than x% of questions correct / total number of pupils, expressed as a percentage	Standard 3 pupils	Non-response' was treated as incorrect in the results. Most non-response happened because of instructions in the test to skip questions if a pupil got a fixed number of prior questions incorrect. 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.
Chapter 3 – Background characteristics of pupils, schools and teachers			
Pupil characteristics			
Poverty score	A pupil is considered 'poor' if he/she comes from a household that has a greater than 50% chance of being below the Tanzania national poverty line, and 'richer' otherwise.	Standard 3 pupils	The poverty scorecard was developed using Tanzania's 2007 Household Budget Survey (Schreiner 2013), and estimates the likelihood that a household has expenditure below a given poverty line (in this case the Tanzania national poverty line). The scorecard uses ten questions: How many household members are 17-years-old or younger, Do all children ages 6 to 17 attend school, Can the female head/spouse read and write, What is the main building material of the floor of the main dwelling, What is the main building material of the roof of the main dwelling, How many bicycles, mopeds, motorcycles, tractors, or motor vehicles does your household own, Does your household own any radios or radio cassettes, Does your household own any lanterns, Does your household own any irons (charcoal or electric) and How many tables does your household own.
Language spoken at home	Indicator variable that equals zero if a pupil speaks Kiswahili at home and equals one if a pupil speaks a local language other than Kiswahili at home.	Standard 3 pupils	The zero category includes a very small proportion of pupils (0.009%) that speaks a non-local language (e.g., English) at home.
Teacher and head teacher background characteristics			
Approaching retirement age 60 (%)	Teachers/head teachers were identified as approaching retirement age if they were at least 56 years old at the time of the survey	Standard 1-3 teachers / head teachers	-
Chapter 4 – School leadership and management			
Head teacher absenteeism			

Head teachers absent on day of survey using head count observation (%)	The number of head teachers who were not present at school on the day of the head count is expressed as a percentage of all head teachers.	Head count	A head count of all teachers including the head teacher was conducted by the enumerator on one of the days they visited the schools.
Head teachers absent "today" using school roster records (%)	The frequency with which the head teachers are absent "today" as per the school records on teacher attendance is expressed as a percentage of all head teachers.	Teacher roster attendance records	Teacher roster records were consulted as a source of information on absenteeism where enumerators recorded whether or not a head teacher was present "today" or each of the previous five days. Teacher roster records may be incomplete but there is no available data or records against which to cross check whether the records are up to date. If head teachers fail to mark their own attendance on the register this measure will identify them as being absent.
Number of days head teacher is absent in past five school days[1]	Head teacher absenteeism for each of the previous five days in the teacher attendance register is summed together. The mean number of days head teachers are absent is reported.		
Absent none of previous 5 days – roster records (%)	Using school teacher roster attendance records, the frequency with which a head teacher was absent over the previous 5 days can be identified. The number of head teachers absent for 'X' out of the previous 5 days, is expressed as a percentage of all head teachers in treatment schools.		
Absent 1 of the previous 5 days – roster records (%)			
Absent 2 of the previous 5 days – roster records (%)			
Absent 3 of the previous 5 days – roster records (%)			
Absent 4 of 5 the previous days – roster records (%)			
Absent 5 of the previous 5 days – roster records (%)			
Whole school development plans and financial management systems			
No whole school development plans (WSDP) for the current year (%)	For each of the indicators, the number of treatment schools in each category is expressed as a percentage of all treatment schools (N=99). Estimates across each category should sum to 100%.	Head teachers	Head teachers were questioned about whether they had a WSDP for the 2014 year. 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, inter alia, teaching and learning objectives, a budget and baseline data and objectives.
Has WSDP for the current year but not available to enumerator (%)			
Has WSDP for the current year and available to enumerator (%)			There is one school with missing data where the head teacher refused to answer the question "Do you have a Whole School Development Plan for the current school year 2014?"
Has WSDP with one of following three elements: budget, teaching and learning objectives, baseline data and targets (%)			

Has WSDP with two of the three elements above (%)			
WSDPs that include all three elements above (%)			
WSDP has a budget, specifically (%)	Number of treatment schools with a WSDP and a budget in the WSDP expressed as a percentage of all treatment schools (except one with missing data on WSDPs).		
Schools that have complete records on per capita grants received in 2012 and 2013 (%)	The number of treatment schools that have complete records on per capita grants received expressed as a percentage of schools that reported receiving capitation grant payments (N=96).	Head teachers	N=96. Among treatment schools, 2 schools reported that they didn't receive capitation grants. Among those that did receive capitation grants 1 head teacher refused to respond the question on complete records and another responded 'don't know'.
Schools that have detailed records of expenditure on the most recent per capita grant received (%)	The number of treatment schools that have detailed records of expenditure on the most recent per capita grant received expressed as a percentage of schools that reported receiving capitation grant payments *and* reported that these payments were made in cash or into the school bank account.	Head teachers	For schools that reported receiving capitation grant payments and reported that these payments were made in cash or into the school bank account, head teachers were asked the following question: "For the last capitation grant payment, is there a record of expenditure available with a breakdown by items purchased" and interviewers proceeded to ask to see a record of expenditure for the last capitation grant payment. The indicator is expressed a percentage of schools receiving capitation grants as cash or into their bank accounts (N=96).
Most commonly used factors to assess teacher performance as reported by head teachers			
Pupil academic results (%)	Head teachers were asked about the most important factor used to assess teacher performance in their school. The number of head teachers choosing each category is expressed as a percentage of head teachers responding to the question. Estimates across each category should sum to 100%.	Head teachers	Data is missing for head teachers of 9 treatment schools. 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, some non-interviewed *head* teachers were phoned for this information to reduce the number of missing observations.
Lesson preparations (%)			
Teaching performance in class (%)			
Teacher punctuality and attendance (%)			
Use of continuous pupil assessment (%)			
Other (%)			
Teacher assessment practices reported by teachers			
Standard 1-3 teachers who report lesson observation by the head teacher in the last 30 days (%)	The number of standard 1-3 teachers who report that the head teacher observed them teaching at least once in the last 30 days expressed as a percentage of standard 1-3 teachers	Standard 1-3 teachers	These indicators related to lesson observations are based on the following questions asked of standard 1-3 teachers in the teacher interview.

Standard 1-3 teachers who *report* written feedback on lesson observation by the head teacher in the last 30 days (%)	The number of standard 1-3 teachers who report that the head teacher observed them teaching and who report that written feedback on the observation was given at least once in the last 30 days; expressed as a percentage of standard 1-3 teachers		• Does your Head Teacher observe your teaching? • How many times did the Head Teacher observe your teaching during the last 30 days? • Do you receive any written feedback from the Head Teacher observations of your teaching? • How many times did you receive written feedback on your teaching during the last 30 days? • Is written feedback from a lesson observation which took place in the last 30 days available?
Standard 1-3 teachers who *show* written feedback on lesson observation by the head teacher in the last 30 days (%)	The number of standard 1-3 teachers who report that the head teacher observed them teaching and who can show the interviewer an example of this written feedback in the last 30 days; expressed as a percentage of standard 1-3 teachers		Two teachers in treatment schools refused to respond to the question "Does your Head Teacher observe your teaching?" reducing the estimation sample to N = 327
Standard 1-3 teachers who report that lesson plans are checked by head teacher or academic master in last 30 days (%)	The number of standard 1-3 teachers who report that the head teacher checks their lesson plans at least once in the last 30 days expressed as a percentage of standard 1-3 teachers	Standard 1-3 teachers	These indicators related to checking lesson plan are based on the following questions asked of standard 1-3 teachers in the teacher interview. • Does your Head Teacher or Academic master check your lesson plans? • How many times did the Head teacher or Academic master check your lesson plans during the last 30 days? • Do you receive any written feedback on your lesson planning from the Head teacher or Academic master? • How many times did you receive written feedback on your lesson plans during the last 30 days? • Is written feedback from lesson plans from the last 30 days available?
Standard 1-3 teachers who *report* written feedback on lesson plans in the last 30 days (%)	The number of standard 1-3 teachers who report that the head teacher checks their lesson plans *and* who *report* that written feedback on these lesson plans was given at least once in the last 30 days; expressed as a percentage of standard 1-3 teachers		
Standard 1-3 teachers who *show* written feedback on lesson plans in the last 30 days (%)	The number of standard 1-3 teachers who report that the head teacher checks their lesson plans *and* who can *show* the interviewer an example of written feedback in the last 30 days; expressed as a percentage of standard 1-3 teachers		
Proportion of standard 1-3 teachers who report receiving at least one performance appraisal in 2013	The number of standard 1-3 teachers who identify that the head teacher, assistant head teacher or academic master holds individual meetings with them to discuss their teacher performance and professional development needs and this happened at least once during 2013, expressed as percentage of standard 1-3 teachers.	Standard 1-3 teachers	This indicator combines responses from two questions "Does the Head Teacher, assistant Head Teacher or academic master hold individual meetings with you to discuss your teacher performance and professional development needs" and "in the calendar year 2013, how many times did you attend an individual professional development meeting?"
Rewards for teaching performance			

Head teacher rewards teachers for performing well (%)	The number of head teachers identifying that there are rewards in their school for teachers who do well; expressed as a percentage of head teachers responding to the question.	Head teachers	Data is missing for head teachers of 10 treatment schools. 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 observations.
Staff meetings			
Head teacher response: 4 or more staff meetings in the last 60 days (%)	The number of head teachers responding that at least 4 staff meetings were held in the last 60 days; expressed as a percentage of head teachers responding to the question.	Head teacher	Data is missing for head teachers of 8 treatment schools. 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.
Teacher response: 4 or more staff meetings in the last 60 days (%)	The number of teachers responding that at least 4 staff meetings were held in the last 60 days; expressed as a percentage of teachers responding to the question.	Standard1-3 teachers	-
Teacher ranking of school support given for improving their teaching			
Very poor (%) Poor (%) Neither poor nor good (%) Good (%) Very good (%)	The number of teachers choosing each ranking on a scale of 1 to 5 is expressed as a percentage of teachers responding to the question. Estimates across each category should sum to 100%.		N = 328. 1 teacher in a treatment school responded "don't know" to the question on the how they would rank the support given by the school to improve their teaching.
Instructional time (see under chapter 5 below)			
Chapter 5 – Teacher professional capacity and performance			
Teacher subject knowledge			
Teacher development needs assessment (TDNA) score (%)	Number of questions answered correctly / total number of questions, expressed as a percentage, for each teacher.	Teachers of standards 1-3 Kiswahili Teachers of standard 1-3 mathematics Teachers of standard 4-7 mathematics	Non-response was treated as incorrect in the teacher score results (person scores) presented in the main body of this report. Non-response rates were fairly high for some of the questions, particularly those which appear later in the TDNA. For this reason, a TDNA score was computed for each question (item scores), but this

		Teachers of standards 1-3 Kiswahili	calculation excludes non-responses if they appear after the last question attempted.
Teacher development needs assessment (TDNA) score for selected topic (%)	Number of questions on topic answered correctly / total number of questions asked on the topic, expressed as a percentage, for each teacher.	Teachers of standard 1-3 mathematics Teachers of standard 4-7 mathematics	
Teacher pedagogy			
Gender balance in teacher interaction with pupils	The number of lessons where teachers' interaction with pupils is gender balanced/all lessons observed, expressed as a percentage (%).	Standard 2 lessons observed	Collection of information: enumerators observed the entire duration of each lesson (a lesson is normally 30, 35 or 40 minutes long), and recorded which pupils teachers interacted with noting whether the pupil was a boy or girl, and how many boys and girls respectively were present. Indicator construction: first, teacher interactions with girls as a proportion of total teacher interactions with pupils (boys and girls) 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.
Spatial balance of teacher interaction with pupils	The number of lessons where the teacher engaged with at least one pupil from all x areas in the classroom/all 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 the enumerator to record the number of interactions between teachers and pupils across the six classroom areas.
Teaching behaviours demonstrated during lessons	The number of lessons where teachers demonstrated fully at least 7 or 3 respectively out of 13 selected teacher skills over the number of all lessons observed, expressed as a percentage (%).	Standard 2 lessons observed	Introductory and concluding lesson stages: for each teaching skill the enumerators recorded responses as follows: 'no' if they did not observe the skills, 'partly' if they observed some of parts of the skill and 'yes' if they observed all of the required aspects of the skill. Middle stage of lesson: for each teaching skill the enumerators recorded responses as follows: 'no' if behaviour not observed, 'yes, infrequently' if skill partly observed and 'yes, frequently' if the skills was frequently observed.
Teacher assessment of pupils' work	The number of teachers who are able to show that they carried out at least two types of pupil assessment (in the five school days before the	Standard 1-3 teacher	Teacher assessment of pupils' work information is obtained from teacher interviews. Teachers were asked what types of pupil assessments they have used in the five school days prior to the interview and to show a

	interview) as a proportion of all teachers interviewed, expressed as a percentage (%).		marked example of each type of assessment they mentioned.
Instructional time			
Actual total days schools were open in 2013	The number of days schools were open in 2013 averaged across treatment schools (N=89).	School records on days open each month in 2013	Total days schools were open in 2013 were identified by summing together total days schools were recorded as open for each month in 2013 using records. This indicator is calculated for 89 schools with no missing records on days open in 2013. There were 11 treatment schools that did not have complete records on days open and it is therefore not possible to construct this indicator for these schools. Among these are also schools where zero days were recorded for some months. Where zeroes were recorded for months other than December and June (school holidays), it was assumed they reflected missing records.
Meets official requirement for open days (%)	The number of schools where recorded open days are 194 days or more, expressed as a percentage of schools with complete open day records.	School records on days open each month in 2013	Schools are officially required to be open for 194 days. As above, N = 89.
Actual weekly timetabled minutes for mathematics in standards 1 and 2 (before adjustment)	Minutes per week timetabled for mathematics in standards 1 and 2 averaged across treatment schools (N=96).	School records – timetables recorded for classes in standards 1 and 2	Data on timetables for each class in standards one and two 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 identify to what extent available instructional time is compromised by absenteeism, indicators on weekly minutes timetabled were then adjusted for whether teachers were present in a classroom. The specific adjustment measure is the number of standard one to three teachers in a school who were present in the class for which they are timetabled to teach in the period before lunch, expressed as a percentage of standard one to three teachers present on the day of survey and are timetabled to teach in the period before lunch. This is a
Meets official requirement for weekly minutes timetabled for mathematics in standards 1 and 2 (*before* adjustment) (%)	The number of schools where minutes calculated per week for mathematics in standard 1 and 2 equals or exceeds an official requirement of 210 minutes per week; expressed as a percentage of treatment schools (N=96).		
Actual weekly timetabled minutes for mathematics in standards 1 and 2 *after* adjusting for the % of teachers timetabled to teach who were present in classroom in the period before lunch.	The minutes per week timetabled for mathematics in standards 1 and 2 after adjusting for the percentage of timetabled standard 1-3 teachers in the classroom in the period before lunch, averaged across treatment schools (N=96).		
Meets official requirement for weekly minutes timetabled for mathematics in standards 1 and 2 (*after* adjustment) (%)	The number of schools where minutes per week calculated for mathematics in standard 1 and 2 (*after* adjustment) equals or exceeds an official requirement of 210 minutes per week; expressed as a % of treatment schools (N=96).		

Actual weekly timetabled minutes for Kiswahili in standards 1 and 2 (*before* adjustment)	Minutes per week timetabled for Kiswahili in standards 1 and 2, averaged across treatment schools (N=92).		rough measure of actual instructional time received. It assumes that the absenteeism behaviour of standards one to three teachers timetabled to teach in the period before lunch is indicative of general teacher behaviour of those teaching Kiswahili and mathematics.
Meets official requirement for weekly minutes timetabled for Kiswahili in standards 1 and 2 (*before* adjustment) (%)	The number of schools where minutes calculated per week for Kiswahili in standard 1 and 2 equals or exceeds an official requirement of 180 minutes per week; expressed as a % of treatment schools (N=92).		Indicators related to mathematics are calculated for treatment schools with complete timetables and non-missing data on the % of standard1-3 teachers timetabled to teach in the period before lunch that were present in that period. After excluding 2 treatment schools with very high weekly minutes in Kiswahili at 600 minutes, the final sample was 92 treatment schools for the Kiswahili related indicators. Indicators related to mathematics are calculated for 96 treatment schools.
Actual weekly timetabled minutes for Kiswahili in standards 1 and 2 *after* adjusting for the % of teachers timetabled to teach who were present in classroom in the period before lunch.	Minutes per week timetabled for Kiswahili in standards 1 and 2 after adjusting for the % of timetabled standard 1-3 teachers in the classroom in the period before lunch, averaged across treatment schools (N=92).		
Meets official requirement for weekly minutes timetabled for Kiswahili in standards 1 and 2 (*after* adjustment) (%)	The number of schools where minutes per weekly calculated for Kiswahili in standard 1 and 2 (*after* adjustment) equals or exceeds an official requirement of 180 minutes per week; expressed as a % of treatment schools (N=92).		
Teacher morale and performance			
Teacher absent on day of survey	The number of teachers who were not present for the teacher head count, expressed as a percentage of all teachers working at the school.	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 actually in classrooms teaching.
Teacher absent from classroom during timetabled lesson (period before lunch)	The number of teachers who were not present at their timetabled lesson before lunch, expressed as a percentage of all teachers present on the day and timetabled to teach the lesson before lunch.	Teachers in schools' teacher rosters who were scheduled to teach before lunch	
Teachers late on day of survey	The number of teachers who arrived after school is officially supposed to start, expressed as a percentage of all teachers present on the day of the survey.	All teachers present in school on day of survey	Note: 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.
Teachers who left classroom during lesson	The number of teachers who left during their lesson, expressed as a percentage of all teachers who were observed teaching	Standard 2 lessons observed	During lesson observations, enumerators recorded if the teacher left the classroom during the lesson.
Teacher job satisfaction and perception of valuation by others			

Teacher job satisfaction	Mean of self-reported ratings of teachers' job satisfaction today	Standard 1-3 teachers	Teachers were asked to report for each of job satisfaction, community appreciation and head teacher appreciation where they placed themselves on a 10-point scale/ladder where 1 is completely unsatisfied and 10 is extremely satisfied. These are subjective measures that are not directly comparable across teachers (who may interpret the instruments differently) but can be used to assess EQUIP-T programme impact (if any) over time on teachers' self-reported job satisfaction and appreciation by the community and head teacher respectively.
Community appreciation	Mean of teachers' ratings of how valued they feel by the community today.		
Head teacher appreciation	Mean of teachers' ratings of how much they feel their head teacher value them as a teacher today.		
Chapter 6 - Community participation and accountability			
School communication with community using notice boards	The number of schools which use notice boards to display different types of information to the public, as a percentage of all schools.	All schools visited	Plan/financial information: information on the Whole School Development Plan, the school budget or school capitation grants. Academic information: academic results or other information related to teaching and learning. Attendance information: pupil or teacher attendance data. Events information: information on school or community events or meetings.
Teacher reporting to parents	The number of teachers who report individual pupil progress to parents, as a percentage of all teachers interviewed.	Standard 1-3 teachers	This information was collected from teachers during teacher interviews. They were asked for the number of times they reported individual pupil progress to parents in the previous academic year.
School committee exist	The number of schools where the head teacher could show minutes from the last meeting of the school committee, expressed as a percentage of all head teachers interviewed.	Head teachers	During interviews, head teachers are asked if the school has a school committee. If so, they are asked to show minutes from their last meeting.
Parents-teachers group exist	The number of schools where the head teacher could show minutes from the last meeting of the parents-teachers group, expressed as a percentage of all head teachers interviewed.		During interviews, head teachers are asked if the school has a parents-teachers group. If so, they are asked to show minutes from their last meeting.
School committee support to school	The number of head teachers who stated that the contribution/support of the school committee to school is good (4) or very good (5), as a percentage of all head teachers interviewed.	Head teachers	During interviews, head teachers are shown a card that displays a 1-5 scale with 1 as 'very poor' and 5 as 'very good'. They are asked to rate the support of the school committee and the community to their school on this scale.
Community support to school	The number of head teachers who stated that the contribution/support of the community to school is good (4) or very good (5), as a percentage of all head teachers interviewed.		

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	Pupil head count and school records	Enumerators enter all standard 1-3 classes and count the number of pupils present. The difference between this head count and the number of standard 1-3 pupils in the school records are considered the number of absent pupils on the day of survey.
Chapter 7 - District and region education management			
Capitation grant payments received as expected	The number of schools that show capitation grant receipts totalling the per pupil amount head teacher expected for the year	Head teacher and school records	Head teacher reported expected capitation grants reported as per pupil amounts; Total capitation grant amounts received reported by head teachers with supporting financial documentation verified by enumerators; Capitation grant amounts received calculated as per pupil amounts using enrolment numbers from 2014.
WEC visit to school	The number of head teachers who report being visited by WEC, as a percentage	Head teachers	During interviews, head teachers are asked if the WEC visited the school in the previous calendar year.
Frequency of WEC visit to school	The mean number of visits by WEC to school reported by head teacher in the previous calendar year	Head teachers	During interviews, head teachers are asked how many times the WEC visited the school in the previous calendar year.
WEC support to school	The number of head teachers who stated that the contribution/support of the WEC to school is good (4) or very good (5), as a percentage of all head teachers interviewed.	Head teachers	During interviews, head teachers are shown a card that displays a 1-5 scale with 1 as 'very poor' and 5 as 'very good'. They are asked to rate the support of the WEC to their school on this scale.
WEC individual visit with teacher	The number of teachers who report being individually visited by WEC, as a percentage	Standard 1-3 teachers	During interviews, teachers are asked if they were individually visited by WEC in the previous calendar year
Frequency of individual WEC visit with teacher	The mean number of visits by WEC with individual teachers in the previous calendar year	Standard 1-3 teachers	During interviews, teachers are asked how many times the WEC visited them individually in the previous calendar year.
DSI visit to school	The number of head teachers who report being visited by DSI, as a percentage	Head teachers	During interviews, head teachers are asked if the DSI visited the school in the previous calendar year.
DSI individual visit with teacher	The mean number of visits by DSI to school reported by head teacher in the previous calendar year	Head teachers	During interviews, head teachers are asked how many times the DSI visited the school in the previous calendar year.
DSI individual visit with teacher	The number of teachers who report being individually visited by DSI, as a percentage	Standard 1-3 teachers	During interviews, teachers are asked if they were individually visited by DSI in the previous calendar year
Frequency of individual DSI visit with teacher	The mean number of visits by DSI with individual teachers in the previous calendar year	Standard 1-3 teachers	During interviews, teachers are asked how many times the DSI visited them individually in the previous calendar year.
Source: OPM impact evaluation team 2014.			

Annex G   Supplementary quantitative baseline analysis

This Annex contains a selection of supplementary quantitative results to those presented in Volume I, Part B chapter two on pupil learning outcomes, and chapter five on teacher professional capacity, performance, motivation and morale.

G.1   Pupil learning outcomes

As context for the pupil learning results from the IE baseline survey, the figure below shows trends in the PSLE pass rate between 2001 and 2013, based on government data.

Chart 6: Pupils passing PSLE by Grade, 2001 - 2013

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Getting A	0.3	0.4	0.8	1.0	0.5	2.0	0.3	0.3	0.4	0.5	1.0	0.3	1.1
Getting B	7.3	8.6	14.2	18.8	20.3	28.3	11.7	10.5	10.9	12.1	16.2	4.9	9.7
Getting C	21.1	18.2	25.1	28.9	40.9	40.2	42.2	44.2	38.1	40.9	41.1	25.5	39.8
Total % Passed	28.6	27.1	40.1	48.7	61.7	70.5	54.2	55.0	49.4	53.5	58.3	30.7	50.6

Source: URT (BEST data for 2001 to 2013)
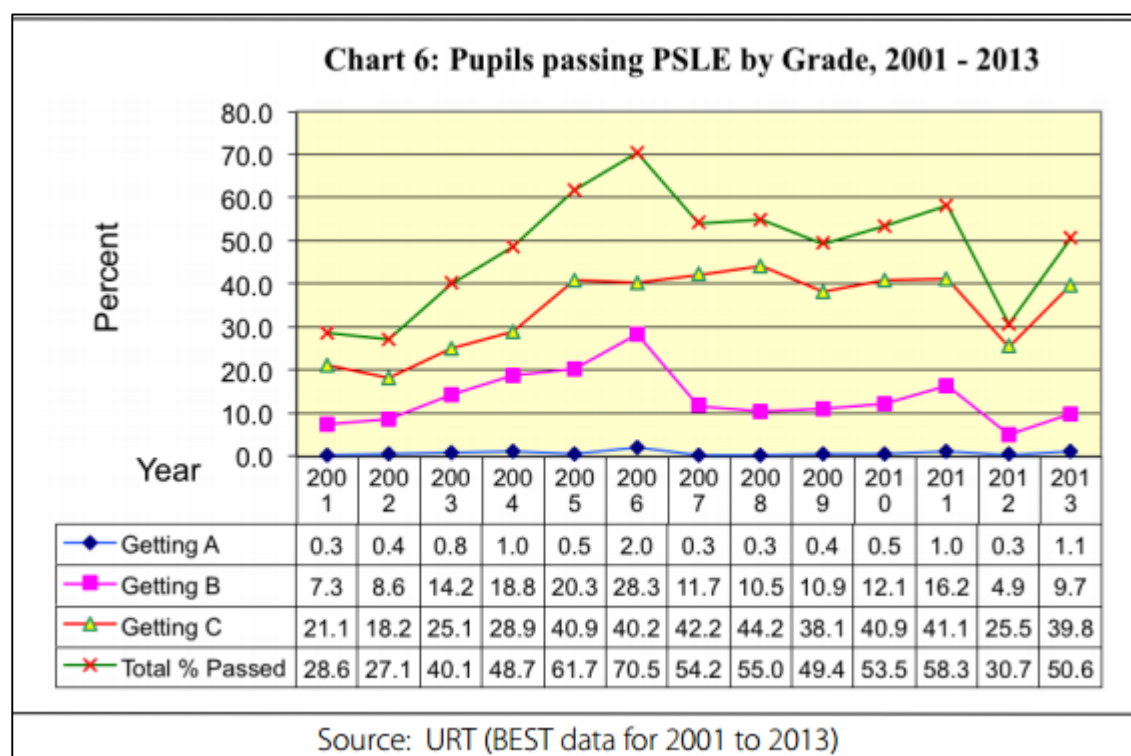
Source: Sumra, S, and Katabaro, J. (2014)

G.1.1 Raw score analysis of pupil Kiswahili and mathematics

The commentary and tables below are based on simple descriptive statistics of pupil test scores from the IE tests. In some tables, an estimate from the 3Rs EGRA or EGMA survey is provided for comparative purposes, although there are serious limitations to this comparison (see Annex box 1 for details).

Annex box 1 Comparability between 3Rs EGRA/EGMA and EQUIP-T IE baseline results

*For information purposes, some of the results from the 3Rs EGRA and EGMA are presented alongside the results from the IE baseline survey. It is important to be cautious in reviewing the two sets of results because they are **not directly comparable for three main reasons**.*

- ***Sample differences**: the 3Rs EGRA/EGMA sample was designed to be nationally representative (RTI 2014, p12) whereas the IE survey results presented in this report are only representative of seventeen EQUIP-T treatment districts, which have significantly different characteristics to the rest of the country). EQUIP-T districts were purposively selected to be among the poorest and most educationally disadvantaged. It is also relevant to note that 77% of pupils in the treatment sample speak a local language at home other than Kiswahili. The national figure, relevant for the 3Rs pupil sample, is likely to be much lower — in UWEZO's 2012 nationally representative learning*

> *assessment roughly one-third of households reported speaking a local language other than Kiswahili at home (UWEZO 2013). On this basis alone, it is reasonable to expect that results will be lower for the programme treatment districts.*
> - ***Differences in groups of tested pupils***: *the 3Rs sampled standard two pupils at the end of the school year, while the IE sampled standard three pupils three months into the school year.*
> - ***Instrument differences***: *the IE and 3Rs instruments are not identical and thus question difficulty may differ, although some component subtests are closer to 'parallel' tests than others.*

PUPIL KISWAHILI RESULTS IN PROGRAMME TREATMENT DISTRICTS

The results are grouped into three skills areas: oral reading, reading and listening comprehension, and writing. For background information on the instruments, including the marking scheme for each subtest and some key measurement issues, see Volume II, Annex I.

Oral reading skills

Standard three pupils are able to read an average of 21 syllables per minute, 14 familiar words per minute, and 9 invented words per minute (Annex table 5). The corresponding estimates from the 3Rs EGRA sample of pupils are markedly higher for reading syllables and familiar words, as anticipated.

As part of a recent target setting exercise for the government's BRN–Education programme, representatives of a wide range of education stakeholders in Tanzania agreed on benchmarks for standard two pupils of an oral reading speed of 50 words per minute, and 40 invented words per minute (RTI 2014, p71). Currently standard three pupils in the programme treatment districts are, on average, not reading accurately at half these benchmark speeds. Another concern, is the range of pupil reading scores. At least 10% of pupils could not read any syllables or words accurately, while pupils close to the top of the distribution of scores (90th percentile) have a reading speed well short of the target benchmarks.

Annex table 5 Pupils' oral reading speed in Kiswahili

Skill area	Indicator	Estimate for treatment pupils	Interdecile range		N (pupils)	3Rs EGRA estimate [1]
			P10	P90		
Syllable sounds	Mean # of correct syllables read per minute	20.9	0.0	49.0	1 491	31.4
Familiar words	Mean # of correct words read per minute	13.7	0.0	34.0	1 496	21.9
Invented words	Mean # of correct words read per minute	9.3	0.0	24.0	1 493	12.3
Reading a passage	Mean # of correct words read per minute	21.3	0.0	54.0	1 496	Not applicable[2]

Sources: IE baseline survey for treatment estimates (pupil Kiswahili test), RTI 2014 for 3Rs. Note: (1) The 3Rs EGMA estimates are provided for information. Results are not directly comparable to the results from the IE baseline survey because of differences in samples and instrument design. (2) For information, the 3R oral reading fluency result was 17.9 words per minute, based on a different passage to the one in the IE test.

Oral reading speed differs significantly between subgroups of pupils divided by poverty status, and by home language, but not by gender (Annex table 6). Pupils who speak Kiswahili at home can on average read eight syllables more per minute than pupils who speak a local language other than Kiswahili at home. The home-language gap ranges from four to nine words per minute for the other tests. Pupils from poorer backgrounds read between three and five fewer syllables or words per minute on average compared with their classmates from richer homes.

Annex table 6 Differences in pupils' oral reading speed in Kiswahili by gender, by poverty status and by home language

Skill area	Indicator	Estimate for treatment pupils[1]					
		Boys	Girls	Poorer	Richer	Home language Kiswahili	Home language other local
Syllable sounds	Mean # of correct syllables read per minute	20.2	21.5	18.6***	22.6	26.8***	19.1
	N (pupils)	(719)	(772)	(477)	(960)	(329)	(1 162)
Familiar words	Mean # of correct words read per minute	13.2	14.3	11.9***	15.1	18.3***	12.4
	N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
Invented words	Mean # of correct words read per minute	8.8	9.8	8.0***	10.3	12.6***	8.3
	N (pupils)	(719)	(774)	(480)	(959)	(330)	(1 163)
Reading a passage	Mean # of correct words read per minute	20.7	22.0	18.3***	23.5	28.2***	19.2
	N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)

Sources: IE baseline survey (pupil Kiswahili test). Note: (1) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Reading comprehension

Pupils were asked five reading comprehension questions based on the short passage they read. The results in Annex table 7 suggest that reading comprehension skills are very weak. There may be various underlying reasons for this, including reading fluency but also poor vocabulary. The average test score was 19%, and more than half of pupils scored zero. In interpreting this result it is important to note that pupils were only asked comprehension questions related to the parts of the passage they were able to read during the timed oral reading test. It is possible that changing the time allowed for oral reading would boost the comprehension scores presented below.

Annex table 7 Pupils' reading comprehension skills in Kiswahili

Skill area	Indicator	Estimate for treatment pupils	Interdecile range		N (pupils)
			P10	P90	
Reading comprehension	Mean test score (%)	19.1	0.0	60.0	1 496
	Percentage of pupils who scored more than 80%	1.2	..	..	1 497
	Percentage of pupils who scored 0%	55.9	..	..	1 496

Source: IE baseline survey (pupil Kiswahili test).

Reading comprehension skills are significantly stronger for some subgroups of pupils then others (Annex table 8). Most likely linked to the findings above on differences in oral reading speed, reading comprehension results are much better for pupils from richer homes compared with those from poorer ones, and are also better for pupils who speak Kiswahili at home compared with their classmates who speak a local language other than Kiswahili at home. There are no significant differences in reading comprehension results by gender.

Annex table 8 Differences in pupils' reading comprehension skills by gender, by poverty status, and by home language

Skill area	Indicator	Estimate for treatment pupils[1]					
		Boys	Girls	Poorer	Richer	Home language Kiswahili	Home language other local
Reading comprehension	Mean test score (%)	18.4	19.6	14.9***	21.8	28.5***	16.2
	N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
	Percentage of pupils who scored more than 80%	1.5	0.9	0.2**	1.7	2.3	0.9
	N (pupils)	(722)	(775)	(481)	(962)	(330)	(1 167)
	Percentage of pupils who scored 0%	57.2	54.7	61.3***	51.8	38.5***	61.2
	N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)

Sources: (i) EQUIP-T IE survey (pupil Kiswahili test). Notes: (1) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Writing skills

Most standard three pupils struggle with standard two curriculum level writing skills: both spelling and punctuation of dictated sentences. Annex table 9 summarises the overall test scores for spelling words, and for punctuating sentences, and presents results for the different subgroups of pupils. Average test scores were 39% for spelling and 30% for punctuation.

Annex table 9 Differences in pupils' writing skills by gender, by poverty status and by home language

Skill area	Indicator	Estimate for treatment pupils						
		All	Boys	Girls	Poorer	Richer	Home language Kiswahili	Home language other local
Spelling	Mean test score (%)	39.1	38.0	40.1	35.1**	42.3	49.4***	36.0
	N (pupils)	(1 496)	(721)	(775)	(480)	(962)	(330)	(1 166)
Punctuation	Mean test score (%)	30.0	29.3	30.7	28.2	31.9	37.2**	27.8
	N (pupils)	(1 496)	(721)	(775)	(480)	(962)	(330)	(1 166)

Sources: IE baseline survey (pupil Kiswahili test). Note: (1) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level. The sample sizes for the sub-groups (boys and girls; poorer and richer; and home language Kiswahili or home language other local) vary somewhat due to some non-responses.

Boys and girls scored similarly in the writing task and there is no significant difference in their test scores. Pupils from poorer backgrounds did significantly worse on the spelling task than the groups from richer homes, but there was little difference in punctuation skills. Again the largest performance differences were found for the home-language subgroups. Pupils who speak a local language other than Kiswahili at home found writing sentences, both spelling and punctuating, significantly more difficult than their classmates who speak Kiswahili at home, and the absolute gaps in average test scores are large[18].

---

[18] If these tests were administered to a nationally representative sample, then it is likely that learning gaps between poorer and richer pupils would be much larger.

PUPIL MATHEMATICS RESULTS IN PROGRAMME TREATMENT DISTRICTS

The results are grouped into three skills areas: basic number skills, addition and subtraction, and multiplication and word problems.

Basic number skills

Pupils performed well on the number comparison subtest (Annex table 10), which required them to choose the larger number from a pair of whole numbers (one digit, two digit and three digit pairs). The average score on this subtest was 65%, and more than one in ten pupils scored full marks. This shows that the majority of standard three pupils have a good understanding of the relative size of numbers. The 3Rs report (RTI 2014, p39) explains that 'being able to compare numbers or quantities is a foundational mathematical skill that is critical to effective and efficient problem solving[19]'. Given its importance as a foundation, it is worth highlighting that a group of pupils are still struggling with number comparison. At least 10% of pupils only managed to get one question out of eight correct.

Test performance on 'missing numbers' which requires pupils to fill in missing numbers in sequences was poor in general, with pupils scoring an average of 29% on these questions (Annex table 10). Only 7% of pupils scored more than 60%, while 13% did not get any answers correct. It appears that vast majority of standard three pupils have not yet developed skills of pattern spotting, which go beyond very simple sequences. The sequences in the test range from single digits, and single steps (difference between adjacent numbers), to two and three digit number sequences, with a range of steps.

Annex table 10 Pupils' mathematics skills in number comparison and missing numbers

Skill area	Indicator	Estimate for treatment pupils	Interdecile range		N (pupils)	3Rs EGMA estimate [1]
			P10	P90		
Number comparison[2]	Mean test score (%)	64.6	12.5	100.0	1 495	61.8
Missing numbers in sequences	Mean test score (%)	28.5	0.0	50.0	1 495	26.1
	Percentage of pupils who scored more than 60%	7.3	..	..	1 495	8
	Percentage of pupils who scored 0%	13.1	..	..	1 495	10.9

Sources: IE baseline survey for treatment estimates (pupil mathematics test), RTI 2014 for 3Rs. Note: (1) The 3Rs EGMA estimates are provided for information. Results are not directly comparable to the results from the EQUIP-T survey because of differences in samples and instrument design. (2) Number comparison is termed 'quantity discrimination' in the 3Rs EGMA report.

On both number comparison and finding missing numbers in sequences, there are statistically significant differences between the performances of subgroups of pupils (Annex table 11). Boys outperform girls, and the gender gap ranges from three to six percentage points depending on the indicator. Pupils from poorer households score an average of four percentage points lower on both topics than their classmates from richer households. But the largest gap in average test scores is between pupils who speak Kiswahili at home and those who speak a local language other than

---

[19] It continues with an example: 'being able to compare numbers or quantities is important when estimating the reasonableness of answers to problems'.

Kiswahili at home. For the number comparison test, the 'home language' gap is eleven percentage points, with the pupils who speak Kiswahili at home being the higher performing group.

Annex table 11 Differences in pupils' basic number skills by gender, by poverty status and by home language

Skill area	Indicator	Estimate for treatment pupils[1]					
		Boys	Girls	Poorer	Richer	Home language Kiswahili	Home language other local
Quantity discrimination	Mean test score (%)	67.0***	62.4	62.2*	66.2	72.9***	62.0
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
Missing numbers	Mean test score (%)	30.4***	26.7	26.4**	29.9	33.0**	27.1
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
	Percentage of pupils who scored more than 60%	8.9*	5.8	5.9	8.3	11.1	6.1
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
	Percentage of pupils who scored 0%	10.1**	15.8	14.2	12.0	8.8**	14.4
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)

Sources: IE baseline survey (pupil mathematics test). Note: (1) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Addition and subtraction

Pupils have reasonable basic addition skills, scoring an average of 61% on the level one questions (Annex table 12). Performance on basic subtraction skills is markedly worse, with the average score on level one subtraction questions being only 46%. Level one questions require pupils to carry out one and two-digit numbers sums with answers up to 99. For both sets of level one questions, the range of performance between the top and bottom 10% of pupils stretches from getting no questions correct to getting full marks.

Moving to level two questions, which include some two- and three-digit sums with answers up to 1000, average test scores plummet to 30% for addition and 20% for subtraction. Clearly the vast majority of standard three pupils have not yet acquired standard-two curriculum level skills in addition and subtraction.

Annex table 12 Pupils' skills in addition and subtraction

Skill area[1]	Indicator	Estimate for treatment pupils	Interdecile range		N (pupils)	3Rs EGMA estimate [2,3]
			P10	P90		
Addition level 1	Mean test score (%)	61.4	0.0	100.0	1 495	71.8%
Addition level 2	Mean test score (%)	30.0	0.0	87.5	1 495	26.1%
Subtraction level 1	Mean test score (%)	45.6	0.0	100.0	1 495	61.1%
Subtraction level 2	Mean test score (%)	19.6	0.0	62.5	1 495	19%
Addition and subtraction level 2	Percentage of pupils who scored more than 80%	7.9	..	..	1 495	8%
	Percentage of pupils who scored 0%	37.8	..	..	1 495	43%

Sources: IE baseline survey for treatment estimate (pupil mathematics test), RTI, Feb 2014 'Baseline Assessment for 3Rs'. Note: (1) Level 1 questions are designed to be easier than level 2 questions. (2) The 3Rs EGMA estimates are

provided for information. Results are not directly comparable to the results from the EQUIP-T survey because of differences in samples and instrument design. (3) The 3Rs results for addition level 1 and subtraction level 2, were based on a timed test, where pupils had 60 seconds to answer as many questions as possible. The result presented is the percentage of sums correct of those attempted.

Similar to the previous findings on basic number skills, Annex table 13 shows that there are significant differences in the performance of different subgroups of pupils on addition and subtraction questions. Strongly significant gender gaps and home language gaps are evident, but the division between performance of poorer and richer pupils is only weakly significant for level two addition questions. Again boys have higher average scores than girls, and the difference is largest (seven percentage points) for the most basic addition and subtraction questions. The size of the advantage in average test scores for pupils who speak Kiswahili at home, ranges from nine to sixteen percentage points.

Annex table 13 Pupil skills in addition and subtraction by gender, by poverty status, and by home language

Skill area	Indicator	Estimate for treatment pupils[1]					
		Boys	Girls	Poorer	Richer	Home language Kiswahili	Home language other local
Addition level 1	Mean test score (%)	64.9***	58.1	60.3	62.8	70.6***	58.5
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
Addition level 2	Mean test score (%)	31.7	28.5	27.1*	32.4	39.9***	27.0
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
Subtraction level 1	Mean test score (%)	49.1***	42.5	44.2	47.4	58.2***	41.8
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
Subtraction level 2	Mean test score (%)	21.1*	18.3	18.8	20.7	26.6**	17.5
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
Addition and subtraction level 2	Percentage of pupils who scored more than 80%	9.3*	6.6	6.6	8.7	13.0*	6.3
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)
	Percentage of pupils who scored 0%	34.7*	40.7	38.5	35.6	26.2***	41.4
	N (pupils)	(721)	(774)	(480)	(961)	(330)	(1 165)

Sources: IE baseline survey (pupil mathematics test). Note: (1) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level. The sample sizes for the sub-groups (boys and girls; poorer and richer; and home language Kiswahili or home language other local) vary somewhat due to some non-responses.

Multiplication and word problems

Most standard three pupils struggle with all but the most basic multiplication sums, and also find word problems difficult. Annex table 14 summarises the average scores for these topics, and presents results for the different subgroups of pupils. Average test scores on multiplication questions were 19%, based on standard two curriculum level questions which require pupils to multiply numbers up to a product of 72. Pupils did better on word problems, but still the average score was only 29%. To answer the word problems, pupils had to solve real-life problems using addition, subtraction and multiplication strategies (one digit numbers).

The most striking feature of the subgroup results on multiplication and word problems is that the home-language gap in average test scores is persistently large and strongly significant (nine and eleven percentage points for the two test topics). This time, boys only outperform girls in word problems and the difference is small (one percentage point) and only weakly significant. Pupils

from poorer background score lower than richer pupils on both skills areas by about four to five percentage points on average.

Annex table 14 Pupil skills in multiplication and word problems by gender, by poverty status and by home language

Skill area	Indicator	Estimate for treatment pupils[1]						
		All	Boys	Girls	Poorer	Richer	Home language Kiswahili	Home language other local
Multiplication	Mean test score (%)	19.4	20.6	18.3	16.7**	21.2	26.5***	17.2
	N (pupils)	(1495)	(721)	(774)	(480)	(961)	(330)	(1 165)
Word problems	Mean test score (%)	28.8	30.1*	27.6	26.5*	30.6	37.5***	26.1
	N (pupils)	(1495)	(721)	(774)	(480)	(961)	(330)	(1 165)

Sources: IE baseline survey (pupil mathematics test). Note: (1) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

G.2 Teacher professional capacity, performance and motivation

G.2.1 Teacher pedagogy

This short section presents two supplementary results on teacher pedagogy observed by enumerators through the lesson observation in programme treatment schools.

Teaching behaviours during the middle stages of the lesson

When it came to the use of teaching aides, less than half of all teachers (44%) used different instructional materials and the majority did so infrequently. By contrast, almost all teachers were found to make effective use of the blackboard (50% did this frequently and 43% infrequently).

Large class sizes and teaching behaviours

It is important to keep in mind that class size affects what teaching behaviour and skills are feasible: large classes generally provide fewer opportunities for using interactive methods and require whole class teaching methods. Large classes are defined here as classes with more than 40 pupils present, the benchmark for teacher to pupil ratio in non-inclusive primary classes set by the government (MOEVT 2009). These large classes present a challenge to teachers. Teachers in large classes were significantly less likely to be observed using paired or group work in class, asking pupils open-ended questions and giving feedback on individual pupil (Annex figure 1).

Annex figure 1 Teaching behaviour and class size

Teaching behaviours influenced by class size

Source: IE Baseline Survey
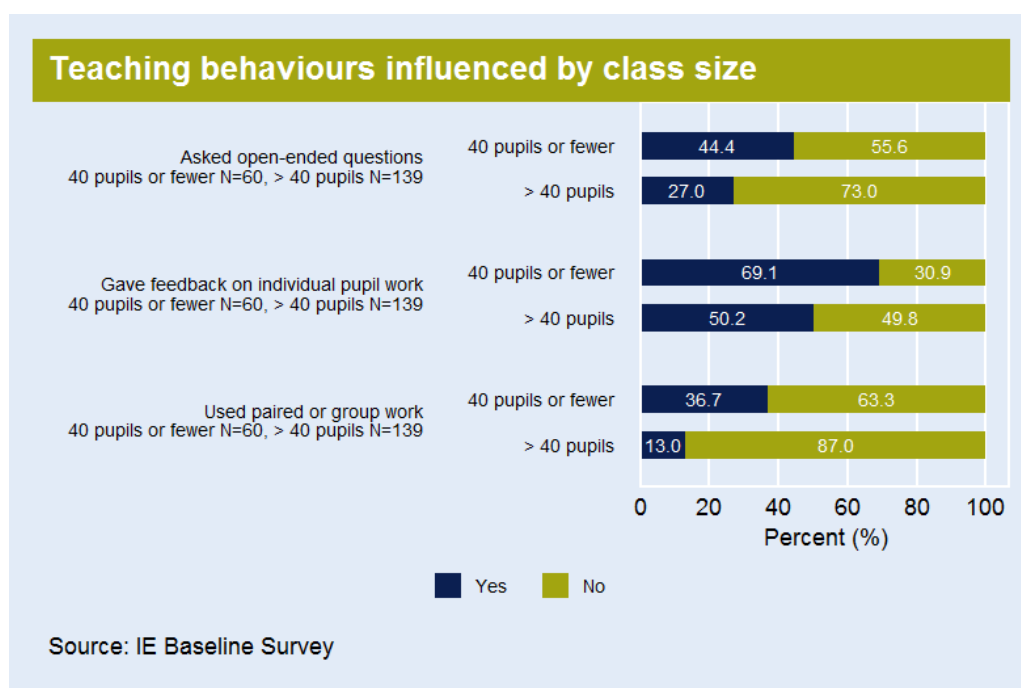
Notes: (1) Weighted estimates.

G.2.2 Teacher subject knowledge and motivation for male and female teachers

This supplementary section discusses key indicators of teacher subject knowledge and teacher motivation disaggregated by gender.

Annex table 15 Teacher subject knowledge in Kiswahili and mathematics by gender

Indicator	Treatment estimates[1]	
	Male	**Female**
TEACHER SUBJECT KNOWLEDGE		
Kiswahili TDNA results (stds 1-3 teachers, intended N up to 300)		
Questions correct (%)	57.8	58.5
N (teachers)	(98)	(149)
Std 1-4 questions correct (%)	65.6	67.0
N (teachers)	(98)	(149)
Std 5-7 questions correct (%)	50.4	50.4
N (teachers)	(98)	(149)
Mathematics TDNA results (stds 1-7 teachers, intended N up to 600)		
Questions correct (%)	64.1***	51.3
N (teachers)	(290)	(216)
Std 1-3 questions correct (%)	89.5	86.1
N (teachers)	(290)	(216)
Std 4-5 questions correct (%)	68.2***	53.3
N (teachers)	(290)	(216)
Std 6-7 questions correct (%)	59.7***	46.5
N (teachers)	(290)	(216)

Teacher subject knowledge in Kiswahili is similar between male and female teachers, according to the TDNA results shown in Annex table 15. For questions testing both lower- and upper-primary level knowledge in Kiswahili, the scores for male and female teachers are almost the same, and there are no significant differences. There is, however, a sizeable gender gap for teacher subject knowledge in mathematics (see lower part of Annex table 15). Overall male teachers outperformed female teachers by 13 percentage points on the total test score, and this difference is strongly significant. This gender gap is driven by differences in performance on the middle- and upper-primary level questions where male teachers scored much better than their female colleagues. On the lower-primary questions (standards one to three), there is no significant gender difference.

Annex table 16 Teacher absenteeism and punctuality by gender

Indicator	Treatment estimates[1]	
	Male	**Female**
TEACHER ABSENCE AND PUNCTUALITY		
Teacher absent from school (%) [all teachers in roster]	12.3	11.8
N (teachers)	(467)	(538)
Teacher arrived late at school (%) [teachers present]	62.3	64.2
N (teachers)	(405)	(468)
Teacher absent from class (%) [teachers present and scheduled to teach]	66.1	67.5
N (teachers)	(323)	(385)

Sources: IE baseline survey. Note: (1) Weighted estimates; (2) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

There are no significant gender differences in levels of teacher absenteeism or punctuality as Annex table 16 reveals. Estimates of rates of school absence, lateness, and classroom absence for male and female teachers are very similar, which gives some indication that levels of motivation among male and female teachers may also be similar.

Annex H   Detailed statistical tables of results from programme treatment districts

This Annex contains detailed statistical tables of baseline results from programme treatment districts. These were presented in summary form in Volume I Part B.

H.1   Pupil learning

Annex table 17 Chapter 3 Pupil learning outcomes

Indicator	Programme treatment districts						
	Mean estimate	Interdecile range		Standard Error	95% confidence interval		N (pupils)
		First decile	Ninth decile		Lower limit	Upper limit	
RAW TEST SCORE RESULTS							
Pupil Kiswahili Test Results (intended N=1,500)							
Syllables/min	20.9	0.0	49.0	1.2	18.5	23.2	1 491
Familiar words/min	13.7	0.0	34.0	0.9	12.0	15.5	1 496
Invented words/min	9.3	0.0	24.0	0.6	8.1	10.5	1 493
Story words/min	21.3	0.0	54.0	1.3	18.8	23.9	1 496
Readcomp qns correct(%)	19.1	0.0	60.0	1.5	16.0	22.1	1 496
Readcomp morethan80%(%)	1.2	..	..	0.4	0.4	1.9	1 497
Readcomp no qns correct(%)	55.9	..	..	2.8	50.3	61.4	1 496
Listcomp qns correct(%)	31.9	0.0	80.0	1.4	29.0	34.7	1 496
Spelling qns correct(%)	39.1	0.0	92.3	2.3	34.5	43.7	1 496
Punctuation qns correct(%)	30.0	0.0	75.0	1.6	26.8	33.3	1 496
Pupil Mathematics Test Results (intended N=1,500)							
QD questions correct(%)	64.6	12.5	100.0	1.6	61.4	67.7	1 495
MissN questions correct(%)	28.5	0.0	50.0	1.1	26.2	30.8	1 495
MissN morethan60% correct(%)	7.3	0.0	0.0	1.3	4.7	9.9	1 495
MissN 0%correct(%)	13.1	0.0	100.0	1.5	10.0	16.1	1 495
add1 questions correct(%)	61.4	0.0	100.0	1.8	57.7	65.0	1 495
add2 questions correct(%)	30.0	0.0	87.5	1.9	26.1	33.9	1 495

sub1 questions correct(%)	45.6	0.0	100.0	1.9	41.9	49.4	1 495
sub2 questions correct(%)	19.6	0.0	62.5	1.7	16.2	23.0	1 495
addsub2_morethan80% correct(%)	7.9	..	..	1.6	4.8	11.0	1 495
addsub2_0%correct(%)	37.8	..	..	2.5	32.9	42.8	1 495
mult questions correct(%)	19.4	0.0	50.0	1.2	16.9	21.8	1 495
word questions correct(%)	28.8	0.0	75.0	1.5	25.7	31.9	1 495
RASCH ANALYSIS RESULTS (intended N=1,500)							
Pupil Kiswahili							
Pupil ability score (logits)	-1.1	-2.9	0.4	0.1	-1.2	-0.9	1 496
pupils in band 0 (%)	38.3	..	..	2.8	32.7	43.9	1 496
pupils in band 1E(%)	8.7	..	..	1.0	6.7	10.6	1 496
pupils in band 1A(%)	16.8	..	..	1.2	14.4	19.3	1 496
pupils in band 2E(%)	24.0	..	..	1.9	20.1	27.9	1 496
pupils in band 2A(%)	12.2	..	..	1.6	9.1	15.4	1 496
Pupil Mathematics							
Pupil ability score (logits)	-0.6	-2.9	1.6	0.1	-0.9	-0.4	1 487
pupils in band 0 (%)	7.4	0.0	0.0	1.0	5.4	9.3	1 487
pupils in band 1E(%)	26.7	0.0	100.0	1.8	23.0	30.3	1 487
pupils in band 1A(%)	30.1	0.0	100.0	1.7	26.8	33.4	1 487
pupils in band 2E(%)	29.7	0.0	100.0	1.9	25.9	33.6	1 487
pupils in band 2A(%)	6.1	0.0	0.0	1.1	3.9	8.3	1 487
Sources: IE baseline survey (pupil tests). Note: (1) Weighted estimates.							

H.2   Supplementary pupil learning analysis

Annex table 18 Chapter 3 Pupil learning outcomes supplementary analysis

Indicator	Estimates for pupils in programme treatment districts[1]					
	Boys	**Girls**	**Poorer**	**Richer**	**Kiswahili**	**Other local Language**
RASCH ANALYSIS RESULTS (intended N=1,500)						

Pupil Kiswahili						
Pupil ability score	-1.1	-1.1	-1.2***	-1.0	-0.7***	-1.2
N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
pupils in band 0 (%)	37.3	39.2	42.2**	35.1	25.7***	42.2
N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
pupils in band 1E(%)	10.5*	7.0	8.5	8.4	7.8	8.9
N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
pupils in band 1A(%)	18.9	15.0	19.4	15.7	18.6	16.3
N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
pupils in band 2E(%)	22.2	25.7	21.6	26.1	28.5	22.6
N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
pupils in band 2A(%)	11.2	13.2	8.3***	14.7	19.5**	10.0
N (pupils)	(721)	(775)	(480)	(962)	(330)	(1 166)
Pupil Mathematics						
Pupil ability score	-0.5***	-0.8	-0.8*	-0.5	-0.0***	-0.8
N (pupils)	(719)	(768)	(477)	(957)	(329)	(1 158)
pupils in band 0 (%)	5.5*	9.1	8.8	6.4	2.2***	9.0
N (pupils)	(719)	(768)	(477)	(957)	(329)	(1 158)
pupils in band 1E(%)	24.0	29.1	27.4	25.6	20.6**	28.5
N (pupils)	(719)	(768)	(477)	(957)	(329)	(1 158)
pupils in band 1A(%)	32.8	27.6	30.9	29.8	29.5	30.3
N (pupils)	(719)	(768)	(477)	(957)	(329)	(1 158)
pupils in band 2E(%)	29.2	30.2	28.6	31.0	37.0**	27.5
N (pupils)	(719)	(768)	(477)	(957)	(329)	(1 158)
pupils in band 2A(%)	8.4***	4.1	4.3*	7.2	10.7*	4.7
N (pupils)	(719)	(768)	(477)	(957)	(329)	(1 158)

Sources: IE baseline survey (pupil tests). Note: (1) Weighted estimates; (2) Statistically significant differences between groups are marked with asterisks on the estimate for the first category in the group: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

H.3 Context pupils, communities and schools

H.3.1 Pupil characteristics

Annex table 19 Chapter 4 Pupil characteristics

| Indicator | Programme treatment districts | | | | | | |
| | Mean estimate | Interdecile range | | Standard Error | 95% confidence interval | | N (pupils) |
		First decile	Ninth decile		Lower limit	Upper limit	
Personal characteristics of pupils (intended N=1,500)							
Girls (%)	52.4	0.0	100.0	1.3	49.8	54.9	1,497
Age (years)	10.3	9.0	12.0	0.1	10.1	10.4	1,248
Household background characteristics (intended N=1,500)							
Belong to a household below poverty benchmark (%)	33.1	0.0	100.0	2.2	28.6	37.6	1,443
Main language spoken at home is a local language, other than Kiswahili (%)	76.6	0.0	100.0	3.0	70.5	82.6	1,497
Sources: IE baseline survey (teacher interview). Note: (1) Weighted estimates.							

H.3.2 School characteristics

Annex table 20 Chapter 4 School characteristics

| Indicator | Programme treatment districts | | | | | | |
| | Mean estimate | Interdecile range | | Standard Error | 95% confidence interval | | N (schools) |
		First decile	Ninth decile		Lower limit	Upper limit	
Pupils[1], teachers[2] and classrooms (intended N=100)							
Number of pupils per school	486.4	233.0	836.0	21.8	442.8	530.1	100
Number of pupils per teacher	53.9	30.8	82.6	2.6	48.7	59.1	100
Number of pupils per class[3] ('class size')	62.7	33.3	92.9	2.7	57.3	68.1	99
Shifting							
Schools where at least one class comes in a different shift (%)	47.9	-	-	6.0	35.9	59.8	100
Water and sanitation (intended N=100)							
Schools with functional toilet (%)	95.7	-	-	0.8	94.2	97.3	99

- Among which, number of pupils per toilet	74.3	26.9	131.4	5.8	62.6	85.9	95
Schools with functional toilet for girls (%)	95.7	-	-	0.8	94.2	97.3	99
- Among which, number of girls per toilet for girls	72.0	26.0	123.0	5.9	60.1	83.8	95
School which have drinking water available on the day of the survey (%)	31.9	-	-	5.2	21.6	42.3	100
Other school resources (intended N=100)							
Schools which have:							
source of electricity which was functioning on the day of the survey (%)	4.5	-	-	2.4	-0.2	9.3	100
separate room where teachers can work (%)	86.1	-	-	4.0	78.1	94.1	100
school library (%)	12.5	-	-	2.4	7.7	17.4	99
at least one working computer (%)	0.9	-	-	1.0	-1.2	2.9	100

Source: IE Baseline Survey. Notes: (1) Unless otherwise stated, indicators refer to pupils/classes in standards one to seven. (2) Includes head teacher. (3) A class is defined as a group of pupils who are taught together.

H.3.3 Head teacher and teacher characteristics

Annex table 21 Chapter 4 Head teacher and teacher characteristics

Indicator	Programme treatment districts						
	Mean estimate	Interdecile range		Standard Error	95% confidence interval		N
		First decile	Ninth decile		Lower limit	Upper limit	
PERSONAL CHARACTERISTICS							
Head teachers (intended N=100)							
Female (%)	15.6	-	-	4.7	6.3	24.9	100
Age (years)	43.4	31.0	56.0	1.1	41.2	45.5	100
Approaching retirement age 60 (%)	14.9	-	-	4.0	6.9	22.9	100
Teachers (intended N up to 300)							
Female (%)	55.2	-	-	3.9	47.4	62.9	329
Age (years)	39.6	24.0	57.0	1.0	37.6	41.7	329
Approaching retirement age 60 (%)	18.5	-	-	2.8	12.7	24.2	329
WORK EXPERIENCE AND TENURE							
Head teachers (intended N=100)							

Time spent as head teacher in current school (years)	3.9	0.0	9.0	0.6	2.7	5.1	100
Time in current school of a year or less (%)	37.5	-	-	6.0	25.5	49.4	100
Teachers (intended N up to 300)							
Time spent as a teacher (years)	15.8	1.0	35.0	1.0	13.8	17.8	329
Time spent as a teacher five years or less (%)	32.1	-	-	3.4	25.3	38.9	329
Time spent as teacher in current school (years)	8.3	1.0	22.0	0.6	7.2	9.5	329
Time in current school of a year or less (%)	18.7	-	-	2.8	13.1	24.2	329
HIGHEST ACADEMIC QUALIFICATION							
Head teachers (intended N=100)							
Certificate or higher (%)	8.6	-	-	3.1	2.3	14.9	100
Form 6 (%)	19.6	-	-	5.5	8.6	30.6	100
Form 4 (%)	71.8	-	-	5.7	60.4	83.1	100
Primary schooling (%)	0.0	-	-	0.0	0.0	0.0	100
Other academic qualification (%)	0.0	-	-	0.0	0.0	0.0	100
Teachers (intended N up to 300)							
Certificate or higher (%)	7.3	-	-	1.8	3.7	10.8	329
Form 6 (%)	3.1	-	-	1.3	0.5	5.8	329
Form 4 (%)	75.8	-	-	3.2	69.4	82.2	329
Primary schooling (%)	13.6	-	-	2.4	8.8	18.4	329
Other academic qualification (%)	0.2	-	-	0.2	-0.2	0.6	329
HIGHEST PROFESSIONAL EDUCATION QUALIFICATION							
Head teachers (intended N=100)							
Diploma in education or higher (%)	9.3	-	-	2.8	3.7	15.0	100
Certificate in education (%)	89.6	-	-	3.1	83.4	95.8	100
Other professional qualification in education (%)	1.1	-	-	1.3	-1.5	3.6	100
No professional qualification in education (%)	0.0	-	-	0.0	0.0	0.0	100
Teachers (intended N up to 300)							

Diploma in education or higher (%)	1.9	-	-	1.0	-0.2	4.0	328
Certificate in education (%)	94.2	-	-	2.1	90.1	98.3	328
Other professional qualification in education (%)	3.3	-	-	2.0	-0.7	7.3	328
No professional qualification in education (%)	0.5	-	-	0.6	-0.7	1.8	328

Sources: IE baseline survey (teacher interview). Note: (1) Weighted estimates; (2) One teacher reported having a professional qualification but did not specify which one. Head teachers were not asked about their total years' experience, only their tenure in the current school. (2) Teachers/head teachers were identified as approaching retirement age if they were at least 56 years old at the time of the survey.

H.4 School leadership and management

Annex table 22 Chapter 5 School leadership and management

Indicator	Programme treatment districts						
	Mean estimate	Interdecile range		Standard Error	95% confidence interval		N
		First decile	Ninth decile		Lower limit	Upper limit	
HEAD TEACHER ABSENTEEISM (intended N=100)							
Head teachers absent on day of survey using head count observation (%)	16.2	-	-	4.0	8.1	24.2	100
Head teachers absent "today" using roster records (%)	22.6	-	-	4.9	12.8	32.4	100
Number of days head teacher is absent in past five school days using roster records	1.3	-	-	0.2	1.0	1.6	100
Absent none of previous 5 days – roster records (%)	36.5	-	-	6.5	23.5	49.6	100
Absent 1 of the previous 5 days – roster records (%)	29.8	-	-	6.5	16.7	42.8	100
Absent 2 of the previous 5 days – roster records (%)	16.3	-	-	5.2	5.9	26.7	100
Absent 3 of the previous 5 days – roster records (%)	8.0	-	-	2.8	2.3	13.6	100
Absent 4 of the previous 5 days – roster records (%)	4.5	-	-	3.0	-1.5	10.4	100
Absent 5 of the previous 5 days – roster records (%)	5.0	-	-	2.7	-0.5	10.4	100
HEAD TEACHER INSET TRAINING ON SCHOOL DEVELOPMENT PLANNING AND MANAGEMENT (intended N=100)							
Any INSET training (%)[2]	11.1	-	-	4.5	2.1	20.2	92
WHOLE SCHOOL PLANNING AND FINANCIAL MANAGEMENT SYSTEMS (intended N=100)							

No whole school development plans (WSDP) for the current year (%)	63.3	-	-	5.3	52.7	73.9	99
Has WSDP for the current year but not available to enumerator (%)	15.9	-	-	4.3	7.3	24.4	99
Has WSDP for the current year and available to enumerator (%)	6.2	-	-	2.4	1.4	11.0	99
Has WSDP with one of following three elements: budget, teaching and learning objectives, baseline data and targets (%)	7.3	-	-	2.6	2.2	12.4	99
Has WSDP with two of the three elements above (%)	5.1	-	-	2.0	1.1	9.0	99
WSDPs that include all three elements above (%)	2.2	-	-	1.6	-0.9	5.4	99
Has WSDP with a budget	9.4	-	-	2.4	4.5	14.3	99
Schools that have complete records on per capita grants received in 2012 and 2013 (%)	83.1	-	-	4.8	73.5	92.7	96
Schools that have detailed records of expenditure on the most recent per capita grant received (%)	83.4	-	-	4.6	74.2	92.6	96
SCHOOL INFORMATION SYSTEMS FOR TEACHER MANAGEMENT							
Most commonly used factors to assess teacher performance as reported by head teachers (intended N=100)[2]							
Pupil academic results (%)	17.7	-	-	5.4	6.9	28.5	91
Lesson preparations (%)	36.5	-	-	6.3	23.9	49.1	91
Teaching performance in class (%)	23.8	-	-	6.4	11.0	36.6	91
Teacher punctuality and attendance (%)	4.1	-	-	2.1	-0.1	8.3	91
Use of continuous pupil assessment (%)	7.6	-	-	2.9	1.7	13.4	91
Other (%)	10.4	-	-	4.8	0.7	20.1	91
Teacher assessment practices reported by teachers (intended N up to 300)							
Standard 1-3 teachers who report lesson observation by the head teacher in the last 30 days (%)	51.5	-	-	4.1	43.3	59.6	327
Standard 1-3 teachers who *report* written feedback on lesson observation by the head teacher in the last 30 days (%)	4.6	-	-	1.5	1.5	7.7	327
Standard 1-3 teachers who *show* written feedback on lesson observation by the head teacher in the last 30 days (%)	1.0	-	-	0.7	-0.3	2.4	327
Standard 1-3 teachers who report that lesson plans are checked by head teacher or academic master in last 30 days (%)	89.4	-	-	2.5	84.4	94.4	329

Standard 1-3 teachers who *report* written feedback on lesson plans in the last 30 days (%)	47.4	-	-	4.2	39.0	55.8	329
Standard 1-3 teachers who *show* written feedback on lesson plans in the last 30 days (%)	33.4	-	-	3.8	25.9	40.9	329
Proportion of standard 1-3 teachers who report receiving a performance appraisal in 2013	27.4	-	-	3.8	19.9	34.9	329
Rewards for teaching performance (intended N=100)[2]							
Head teacher rewards teachers for performing well (%)	34.7	-	-	7.0	20.7	48.8	90
Staff meetings (HT: intended N=100, teachers intended N up to 300)[2]							
Head teacher response: 4 or more staff meetings in the last 60 days (%)	23.2	-	-	5.4	12.2	34.1	92
Teacher response: 4 or more staff meetings in the last 60 days (%)	15.3	-	-	2.9	9.6	21.1	329
Teacher ranking of school support given for improving their teaching (intended N up to 300)							
Very poor (%)	5.1	-	-	1.3	2.5	7.7	328
Poor (%)	2.9	-	-	0.9	1.1	4.7	328
Neither poor nor good (%)	40.9	-	-	3.7	33.5	48.3	328
Good (%)	38.9	-	-	3.7	31.4	46.3	328
Very good (%)	12.3	-	-	2.6	7.1	17.4	328
INSTRUCTIONAL TIME (intended N=100)							
School open days							
Official requirement for school open days	*194.0*	-	-	-	-	-	-
Actual total days schools were open in 2013	193.4	179.0	210.0	1.5	190.3	196.5	89
Meets official requirement for open days (%)	43.6	-	-	5.0	33.6	53.6	89
Weekly timetabled minutes for mathematics (intended N=100)							
Official requirement for weekly timetabled minutes for mathematics in standards 1 and 2	*210.0*	-	-	-	-	-	-
Actual weekly timetabled minutes for mathematics in standards 1 and 2 (before adjustment)	219.3	180.0	285.0	7.0	205.2	233.4	96
Meets official requirement for weekly minutes timetabled for mathematics in standards 1 and 2 (*before* adjustment) (%)	67.8	0.0	100.0	6.8	54.1	81.4	96
Actual weekly timetabled minutes for mathematics in standards 1 and 2 *after* adjusting for the % of teachers	77.0	0.0	195.0	12.0	52.9	101.2	96

timetabled to teach who were present in classroom in the period before lunch.							
Meets official requirement for weekly minutes timetabled for mathematics in standards 1 and 2 (*after* adjustment) (%)	8.7	0.0	0.0	4.2	0.2	17.1	96
Weekly timetabled minutes for Kiswahili (intended N=100)							
Official requirement for weekly timetabled minutes for Kiswahili in standards 1 and 2	*180.0*	-	-	-	-	-	-
Actual weekly timetabled minutes for Kiswahili in standards 1 and 2 (*before* adjustment)	207.1	150.0	330.0	8.6	189.8	224.3	92
Meets official requirement for weekly minutes timetabled for Kiswahili in standards 1 and 2 (*before* adjustment) (%)	82.4	0.0	100.0	5.6	71.2	93.6	92
Actual weekly timetabled minutes for Kiswahili in standards 1 and 2 *after* adjusting for the % of teachers timetabled to teach who were present in classroom in the period before lunch.	78.4	0.0	180.0	11.0	56.2	100.6	92
Meets official requirement for weekly minutes timetabled for Kiswahili in standards 1 and 2 (*after* adjustment) (%)	10.5	0.0	100.0	4.0	2.4	18.6	92

Sources: IE baseline survey (teacher interview). Note: (1) Weighted estimates; (2) At some schools the head teacher was not present on either day of the survey and instead the deputy head teacher responded, but some questions were designed to be asked of the head teacher only leading to some missing observations.

H.5   Teachers capacity, performance, motivation and morale

Annex table 23 Chapter 6 Teacher subject knowledge, pedagogy, motivation and morale

Indicator	Programme treatment districts						
	Mean estimate	Interdecile range		Standard Error	95% confidence interval		N (pupils)
		First decile	Ninth decile		Lower limit	Upper limit	
TEACHER SUBJECT KNOWLEDGE							
Kiswahili TDNA results (intended N up to 300)							
questions correct_K (%)	58.2	39.5	74.4	1.1	56.0	60.4	249
std1-4 qns correct_K(%)	66.4	42.9	85.7	1.5	63.4	69.4	249
std5-7 qns correct_K(%)	50.4	36.4	68.2	1.0	48.3	52.5	249
reading-comp qns correct(%)	68.3	50.0	80.0	1.1	66.1	70.6	249
gram-punct qns correct(%)	42.1	14.3	71.4	1.7	38.7	45.4	249
dir/ind/tenses qns correct(%)	68.6	44.4	88.9	1.6	65.5	71.8	249

synonyms proverbs qns correct(%)	61.3	30.0	90.0	1.6	58.2	64.5	249
Mathematics TDNA results (stds 1-7 teachers, intended N up to 600)							
qns correct_M(%)	59.0	27.8	86.1	1.3	56.5	61.6	506
std1-3 qns correct_M(%)	88.1	66.7	100.0	1.1	85.9	90.4	506
std4-5 qns correct_M(%)	62.3	25.0	87.5	1.3	59.8	64.9	506
std6-7 qns correct(%)	54.5	20.0	88.0	1.6	51.4	57.6	506
wholeNo qns correct(%)	69.0	42.9	85.7	1.2	66.6	71.4	506
Frac/dec/perc qns correct(%)	66.7	27.3	90.9	1.1	64.6	68.9	506
Geometry qns correct(%)	43.1	0.0	83.3	2.0	39.2	47.0	506
Stats qns correct(%)	65.9	20.0	100.0	1.9	62.1	69.7	506
Algebra qns correct(%)	45.4	0.0	100.0	3.0	39.5	51.3	506
TEACHER PEDAGOGY							
Gender balance in teacher-pupil interactions during lessons (intended N=200)							
Teacher interactions were gender balanced (%)	53.8	.	.	4.4	45.1	62.4	193
Teacher interacted with more boys (%)	30.3	.	.	3.8	22.8	37.8	193
Teacher interacted with more girls (%)	16.0	.	.	3.3	9.5	22.4	193
Spatial balance in teacher-pupil interactions during lessons (intended N=200)							
Teacher interacted with all 6 areas (%)	58.5	.	.	4.0	50.6	66.3	199
Teacher interacted with 4 areas (%)	87.2	.	.	2.7	81.9	92.4	199
Interactions with front two areas (mean %)	41.5	23.5	66.7	1.5	38.4	44.5	193
Interactions with middle two areas (mean %)	30.4	8.3	47.6	1.5	27.3	33.4	193
Interactions with back two areas (mean %)	28.2	10.3	45.5	1.2	25.8	30.6	193
TEACHER PEDAGOGY							
Teaching behaviours FREQUENTLY observed during lessons (intended N=200)							
States objectives (%)	23.0	.	.	3.1	16.9	29.2	199
States new skills (%)	7.2	.	.	1.8	3.6	10.8	199
Checks prior knowledge (%)	21.5	.	.	2.8	16.1	27.0	199
Checks acquired new skills (%)	21.8	.	.	2.9	16.1	27.5	199
Plenary summarising learning (%)	21.1	.	.	2.6	16.1	26.2	199
Pupils demonstrate in front of class (%)	35.8	.	.	3.6	28.7	42.9	199

Asks open-ended questions (%)	11.2	.	.	2.3	6.7	15.7	199
Probes pupil answers (%)	11.8	.	.	1.9	8.1	15.5	199
Encourages pupil questions (%)	4.0	.	.	1.8	0.4	7.6	199
Gives feedback on pupil work (%)	25.7	.	.	3.4	18.9	32.6	199
Uses paired or group work (%)	6.5	.	.	1.7	3.2	9.8	199
Makes effective use of blackboard (%)	49.9	.	.	3.4	43.2	56.5	199
Uses different instructional materials (%)	18.3	.	.	3.0	12.4	24.3	199
Relates well with & praises pupils (%)	51.0	.	.	3.6	43.8	58.2	199
Number of teaching behaviours observed during lessons (intended N=200)							
Teacher frequently demonstrates 0-2 behaviours (%)	42.3	.	.	3.5	35.5	49.1	199
Teacher frequently demonstrates 3-6 behaviours (%)	48.7	.	.	3.6	41.7	55.7	199
Teacher frequently demonstrates >6 behaviours (%)	9.0	.	.	1.8	5.4	12.5	199
Teacher frequently or infrequently demonstrates 0-2 behaviours (%)	4.6	.	.	2.0	0.7	8.6	199
Teacher frequently or infrequently demonstrates 3-6 behaviours (%)	28.5	.	.	2.6	23.3	33.7	199
Teacher frequently or infrequently demonstrates >6 behaviours (%)	66.8	.	.	3.0	60.8	72.8	199
Teacher assessment of pupil work (intended N up to 300)							
Teacher could show use of at least two assessment types (%)	26.7	.	.	2.4	22.0	31.5	324
Teacher could not show any marked assessment (%)	30.0	.	.	2.4	25.2	34.8	324
Teacher showed class exercise (%)	53.5	.	.	2.8	48.0	59.1	324
Teacher showed written class test (%)	40.2	.	.	2.8	34.7	45.6	324
Teacher showed homework (%)	4.7	.	.	1.7	1.4	7.9	324
Teacher showed other written assessment (%)	5.7	.	.	1.5	2.7	8.6	324
Teacher showed oral evaluation (%)	1.4	.	.	0.5	0.4	2.4	324
Language used by teachers during lessons (intended =200)							
Of lessons observed:							
Teacher did not switch language / Kiswahili (%)	96.0	.	.	1.4	93.3	98.8	199

Teacher infrequently switched between Kiswahili and local language (%)	4.0	.	.	1.4	1.2	6.7	199
TEACHER MOTIVATION AND MORALE							
School absenteeism (%)	12.1	.	.	1.1	9.9	14.2	1,005
Teachers arriving late (of teachers present) on day of the survey (%)	63.2	.	.	3.8	55.6	70.9	873
Classroom absenteeism (%)	66.8	.	.	3.0	60.8	72.7	708
Teachers who left the lesson (%)	15.3	.	.	3.3	8.8	21.8	199
Job satisfaction (scale of 1 to 10)	7.7	5.0	10.0	0.2	7.4	8.1	329
Valued by community (scale of 1 to 10)	6.5	2.0	10.0	0.3	6.0	7.0	326
Valued by head teacher (scale of 1 to 10)	8.5	6.0	10.0	0.1	8.3	8.8	328

Sources: IE baseline survey (teacher interview). Note: (1) Weighted estimates.

H.6 Supplementary teacher subject knowledge analysis

Annex table 24 Supplementary TDNA Kiswahili analysis: item scores (excluding non-response)

Topic	Assessment item	Estimate for teachers in treatment districts	SE	Lower 95% CI	Upper 95% CI	N (teachers)
Reading Comprehension	qn1 correct(%)	93.1	2.3	88.4	97.8	249
	qn2 correct(%)	93.8	2.3	89.1	98.4	249
	qn3 correct(%)	86.0	3.0	80.0	92.0	249
	qn4 correct(%)	55.2	4.0	47.3	63.1	249
	qn5 correct(%)	69.2	3.6	62.1	76.3	249
	qn6 correct(%)	11.9	2.9	6.1	17.8	249
	qn7 correct(%)	82.8	2.8	77.3	88.3	249
	qn8 correct(%)	21.4	4.2	13.0	29.9	249
	qn9 correct(%)	82.6	3.0	76.6	88.6	249
	qn10 correct(%)	87.4	2.5	82.3	92.5	249
Grammar	qn11 correct(%)	82.2	3.0	76.2	88.2	248
	qn12 correct(%)	54.8	4.7	45.5	64.2	248
	qn13 correct(%)	57.0	4.6	47.7	66.3	247
	qn14 correct(%)	66.6	4.2	58.2	75.0	247

	qn15 correct(%)	78.2	2.9	72.5	83.9	247
	qn16 correct(%)	9.5	2.2	5.2	13.9	247
	qn17 correct(%)	57.9	3.9	50.0	65.8	247
Synonyms	qn18 correct(%)	80.6	2.8	74.9	86.3	245
	qn19 correct(%)	39.0	4.5	29.9	48.0	245
	qn20 correct(%)	43.2	3.5	36.2	50.2	245
	qn21 correct(%)	94.5	1.7	91.0	97.9	245
	qn22 correct(%)	69.7	3.7	62.4	77.1	245
Proverbs	qn23 correct(%)	79.5	3.1	73.3	85.7	245
	qn24 correct(%)	52.6	4.8	43.0	62.3	245
	qn25 correct(%)	93.0	2.3	88.4	97.5	244
	qn26 correct(%)	77.4	4.1	69.2	85.7	244
Direct and indirect speech	qn27 correct(%)	18.5	3.7	11.0	26.0	244
	qn28 correct(%)	81.2	4.8	71.6	90.7	244
	qn29 correct(%)	90.4	2.4	85.5	95.2	244
	qn30 correct(%)	78.1	3.3	71.4	84.8	244
Tenses	qn31 correct(%)	87.2	2.7	81.9	92.6	244
	qn32 correct(%)	37.7	4.4	28.8	46.5	242
	qn33 correct(%)	88.7	2.4	83.8	93.6	241
	qn34 correct(%)	69.8	3.6	62.7	77.0	241
	qn35 correct(%)	12.0	2.6	6.7	17.3	203
	qn36 correct(%)	44.8	3.9	36.9	52.6	203
	qn37 correct(%)	52.0	4.9	42.2	61.9	203
	qn38 correct(%)	46.4	4.5	37.3	55.4	203
Punctuation	qn39 correct(%)	52.2	4.8	42.5	61.9	203
	qn40 correct(%)	71.2	3.7	63.7	78.8	203
	qn41 correct(%)	16.3	3.5	9.3	23.2	203
	qn42 correct(%)	9.2	2.8	3.7	14.7	203
	qn43 correct(%)	7.2	2.4	2.4	12.1	203

Sources: IE baseline survey (TDNA Kiswahili).

Annex table 25 Supplementary TDNA mathematics analysis: item scores (excluding non-response)

		Estimate for teachers in treatment districts	SE	Lower 95% CI	Upper 95% CI	Treatment N
Whole Numbers	qn1 correct(%)	93.4	1.5	90.4	96.4	507
	qn2 correct(%)	77.7	2.2	73.3	82.2	507
	qn3 correct(%)	93.3	1.3	90.8	95.9	507
	qn4 correct(%)	89.8	1.8	86.2	93.4	507
	qn5 correct(%)	66.9	3.1	60.7	73.0	507
	qn6 correct(%)	43.6	3.2	37.2	50.1	507
	qn7 correct(%)	18.0	2.5	13.1	22.9	507
Fractions	qn8 correct(%)	65.6	2.3	61.0	70.3	507
	qn9 correct(%)	88.1	1.8	84.4	91.8	507
	qn10 correct(%)	70.4	2.6	65.1	75.7	507
	qn11 correct(%)	72.9	2.2	68.4	77.4	507
Decimals	qn12 correct(%)	82.5	1.9	78.7	86.2	507
	qn13 correct(%)	51.0	3.3	44.4	57.6	507
	qn14 correct(%)	73.9	3.0	67.9	79.8	507
	qn15 correct(%)	72.2	2.4	67.4	77.0	507
Percentages	qn16 correct(%)	89.3	1.8	85.7	93.0	506
	qn17 correct(%)	64.5	2.7	59.0	70.0	506
	qn18 correct(%)	4.1	1.0	2.1	6.1	506
Measurement	qn19 correct(%)	53.9	2.5	48.8	58.9	506
	qn20 correct(%)	40.6	2.6	35.5	45.7	505
Geometry	qn21 correct(%)	53.3	3.4	46.4	60.2	502
	qn22 correct(%)	61.4	2.8	55.8	67.0	502
	qn23 correct(%)	14.6	2.4	9.8	19.4	502
	qn24 correct(%)	44.8	3.7	37.4	52.2	497
	qn25 correct(%)	45.8	3.3	39.2	52.5	495

	qn26 correct(%)	43.3	3.2	36.9	49.6	494
	qn27 correct(%)	84.7	2.2	80.4	89.1	491
	qn28 correct(%)	43.5	3.5	36.4	50.6	487
Statistics	qn29 correct(%)	81.6	2.1	77.4	85.7	483
	qn30 correct(%)	73.7	2.8	68.2	79.3	480
	qn31 correct(%)	62.6	2.9	56.9	68.3	477
	qn32 correct(%)	56.6	3.9	48.7	64.5	461
	qn33 correct(%)	77.9	3.0	71.9	83.8	454
Algebra	qn34 correct(%)	46.3	3.9	38.6	54.1	434
	qn35 correct(%)	47.8	5.1	37.7	58.0	412
	qn36 correct(%)	42.2	4.9	32.3	52.1	372

Sources: IE baseline survey (TDNA Kiswhili).

H.7 Community participation and accountability

Annex table 26 Chapter 7 Community participation and accountability

Indicator	Programme treatment districts						
	Mean estimate	Interdecile range		Standard Error	95% confidence interval		N
		First decile	Ninth decile		Lower limit	Upper limit	
Communication between school and community (intended N=100)							
Of all schools visited:							
School had notice board (%)	49.0	.	.	6.8	35.4	62.6	100
Of schools with a notice board:							
Notice board displayed school plan, budget or capitation grant information (%)	3.1	.	.	2.7	-2.7	8.9	52
Notice board displayed academic progress or other information about teaching and learning (%)	21.1	.	.	9.8	0.3	41.8	52
Notice board displayed pupil and teacher attendance information (%)	16.8	.	.	5.0	6.2	27.5	52
Notice board displayed information on school or community meetings or events (%)	15.8	.	.	5.9	3.3	28.3	52
Of teachers interviewed: (intended N up to 300)							

Teacher reported individual pupil progress to parents in 2013 (%)	75.5	.	.	2.9	69.7	81.4	329
Teacher reported individual pupil progress to parents more than twice in 2013 (%)	2.9	.	.	1.2	0.4	5.4	254
Capacity and engagement of community (intended N=100)							
Of head teacher interviews:							
School has school committee and head teacher could show minutes from a previous meeting (%)	91.4	.	.	3.8	83.7	99.0	100
Head teacher rating of school committee support to school (1-5 scale)	3.6	3.0	4.0	0.1	3.4	3.7	99
Parent teacher association/group exists (%)	14.0	.	.	4.1	5.7	22.3	100
Parent-teacher meetings held at least yearly (%)	86.7	.	.	3.1	80.5	92.8	100
School has parents-teachers meeting at least yearly and head teacher could show minutes from previous meeting (%)	67.4	.	.	5.2	57.0	77.9	100
Head teacher rating of local community support to school (1-5 scale)	2.7	1.0	4.0	0.1	2.5	3.0	100
Of schools visited:							
From school records: absence rate of standard 1-3 pupils, total (%)	29.6	13.9	49.3	1.4	26.8	32.3	100
From school records: absence rate of standard 1-3 pupils, boys (%)	30.8	14.6	52.5	1.7	27.5	34.2	100
From school records: absence rate of standard 1-3 pupils, girls (%)	28.5	11.1	48.7	1.3	25.9	31.1	100
From enumerator head count: absence rate of standard 1-3 pupils, total (%)	32.9	13.5	53.6	1.6	29.6	36.2	100
From enumerator head count: absence rate of standard 1-3 pupils, boys (%)	34.1	15.1	57.9	1.9	30.3	37.9	100
From enumerator head count: absence rate of standard 1-3 pupils, girls (%)	31.7	11.8	49.1	1.6	28.5	35.0	100

Sources: IE baseline survey (teacher interview). Note: (1) Weighted estimates.

H.8   District and region education management

Annex table 27 Chapter 8 District and region education management

Indicator	Programme treatment districts						
	Mean estimate	Interdecile range		Standard Error	95% confidence interval		N (pupils)
		First decile	Ninth decile		Lower limit	Upper limit	
Financial planning and management (intended N=100)							
Of schools with complete financial records:							
Capitation grant payments per pupil expected in 2012 (Tsh)	8,408.8	1,000.0	10,000.0	885.3	6,605.5	10,212.0	68
Estimate of capitation grant payments per pupil received in 2012 (Tsh)	3,488.8	1,197.3	7,812.8	330.9	2,821.5	4,156.1	85
Amount of capitation grant payments received in 2012 met or exceeded amount expected (%)	11.1	.	.	4.0	3.1	19.1	100
Capitation grant payments per pupil expected in 2013 (Tsh)	8,396.4	1,222.0	10,000.0	888.3	6,587.1	10,205.8	68
Estimate of capitation grant payments per pupil received in 2013 (Tsh)	3,842.2	542.6	7,927.2	604.9	2,624.5	5,059.9	90
Amount of capitation grant payments received in 2013 met or exceeded amount expected (%)	12.2	.	.	5.0	2.1	22.2	100
District performance management							
Of head teachers interviewed (intended N=100)							
School was visited by WEC in 2013 (%)	99.1	.	.	0.0	99.0	99.2	100
Of those visited, number of visits by WEC in 2013	6.6	2.0	12.0	0.7	5.2	8.0	99
Head teacher rating of WEC support (1-5 scale)	4.0	3.0	5.0	0.1	3.8	4.2	100
School was visited by DSI 2013 (%)	63.4	.	.	5.7	52.0	74.7	100
Of those visited, number of visits by DSI in 2013	1.5	1.0	3.0	0.1	1.2	1.9	61
Of teachers interviewed: (intended N up to 300)							
Visited by WEC individually in 2013 (%)	33.5	.	.	3.7	26.1	40.8	329
Of those visited, frequency of WSE visits in 2013	2.7	1.0	5.0	0.4	2.0	3.5	118
Visited by DSI individually in 2013 (%)	22.4	.	.	2.8	16.8	28.0	329
Of those visited, frequency of DSI visits	1.3	1.0	2.0	0.1	1.0	1.6	76

Receipt of in-kind resources in 2012 and 2013 (intended N=100)							
Receives any in-kind resources (%)	87.8	-	-	4.0	79.8	95.8	100
Receives textbooks (%)	80.5	-	-	5.2	70.1	91.0	100
Receives classroom furniture (%)	23.7	-	-	5.1	13.6	33.8	100
Receives classroom/s (%)	7.5	-	-	3.5	0.6	14.5	100
Receives water/toilets (%)	8.7	-	-	3.2	2.3	15.1	100
Receives teachers' houses (%)	6.5	-	-	3.6	-0.7	13.7	100

Sources: IE baseline survey (teacher interview). Note: (1) Weighted estimates. Estimate of capitation grant payments received annually were based on the total amounts received for the relevant year divided by the number of pupils enrolled in 2014 as an indication of school size.

Annex I   Pupil learning outcomes technical annex

I.1   Summary of the content of the pupil tests

I.1.1   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: Reading comprehension is a standard two level curriculum skill. The remaining subtests combine curriculum one and curriculum two level skills by including questions of different levels within each subtest. Reading speed is not specified in the curriculum, but well-known international research (Abadzi, 2006), suggests that 45-60 words per minute is a minimum for reading fluency sufficient for reading comprehension. As part of a target-setting exercise for the government's BRN-education programme in Tanzania, a target was set of 50 familiar words per minute for standard two pupils.

Comparison with contents of 3Rs EGRA: (i) the timed oral reading tests of syllables, familiar words and invented words in the IE test were designed to be similar (close to 'parallel') to the 3Rs EGRA questions, although the actual items are different; (ii) the reading and listening comprehension passages, and the dictated sentences, in the IE test are completely different to the 3Rs EGRA passages and sentences.

I.1.2   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 which is a standard two level competency, the other five subtests contain a mix of standard one and standard two level questions. Over the whole test the balance is skewed towards standard one level material; about 60% of the questions are at the lower curriculum level.

Comparison with contents of 3Rs EGMA: (i) the IE test contains one extra topic/subtest (multiplication) but the rest of the topics are the same; (ii) the IE test contains standard two curriculum level addition and subtraction questions which are not part of 3Rs (both tests contain standard-one curriculum level questions); (iii) the IE test contains slightly fewer questions in most topics/subtests; (iv) apart from the exceptions mentioned already, the IE test questions were designed to be similar (or close to 'parallel') to the 3Rs EGMA questions.

I.2   Notes on traditional test analysis

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. 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 affect the 'reading speed' indicator. Most non-response happened because of instructions in the test to skip questions if 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.

I.3 Rasch analysis

The purpose of this section is to: (i) set out the reasons for using Rasch modelling to analyse the Kiswahili and mathematics test data for the IE baseline; (ii) explain the principles underpinning the Rasch model and some of its key assumptions; (iii) describe the approach taken to the Rasch analysis of the IE test, and present the competency band descriptors; and (iv) discuss the diagnostic tests that were carried out to produce the final estimates that were presented in Volume I of this report.

I.3.1 Overview of Rasch model

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 (i.e. one that cannot be measured directly) that can be thought of in terms of more or less. In the IE, the latent unidimensional traits that we are seeking to measure using the pupil tests are abilities/performance in Kiswahili literacy and mathematical numeracy.

The Rasch model is the simplest latent trait model. It is based on a mathematical model of a person's (in our case, a pupil's) response to an item. 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 Rasch model is probabilistic, and it is special case of an item-response theory (IRT) model. 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 difficulty of item v and the ability of person i

The Rasch model enables the creation of an interval scale of scores for both the item's difficulty and the person's ability, and these scores are scaled in logits. The Rasch model has the property of specific objectivity, which may be interpreted as sample independence. This is its advantage over other IRT models. The Rasch model is easily extended under the same core principle to items with ordered polytomous responses. The IE analysis applies the polytomous Rasch models 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. A summary of the results from the diagnostics tests that were carried out on the Kiswahili and mathematics item responses, are in the final section I.3.2.3. This analysis led to the decision to remove two of the Kiswahili subtests from the final analysis (syllables and listening comprehension), but to keep all of the mathematics subtests with only the removal of one problematic item (from the number comparison subtest).

Source: this text was partly extracted from: Cueto et al. 2009.

I.3.2   Approach to Rasch analysis and performance band competency descriptors

I.3.2.1   Kiswahili
Treatment of items
Each of the three reading subtests included in the analysis (familiar words, non-words, and story passage) is treated as a separate polytomous item, which means that there are more than two answer categories. The answer categories were 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 were included in one answer category. For example, on the familiar words subtest, the answer categories ranged from one word per minute to a top category of 46 words per minute or above.
For the remaining subtests, each is treated as a testlet in the analysis to account for the dependence between them (this is explained later in the diagnostics section). 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.

Treatment of non-response
The treatment of non-response differs by subtest for various reasons:
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—first sentence: non-response is treated as incorrect

Writing—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.

Rasch analysis

The Rasch analysis of the Kiswahili pupil test produces estimates of person ability and item difficulty on an interval scale, summarised in the item location map below. The left-hand side of the scale shows the frequency distribution of the person ability estimates, while the right-hand side shows the location of the items (NB: this is a simplified map, and not all items are shown). The non-normal distribution is discussed later in the diagnostics section.

The following abbreviations are used to label the items from each subtest: **FW:** familiar words **NW:** non-words **ORP**: Oral reading passage **RC:** reading comprehension **Sp**: Writing spelling **Pu**: Writing punctuation.
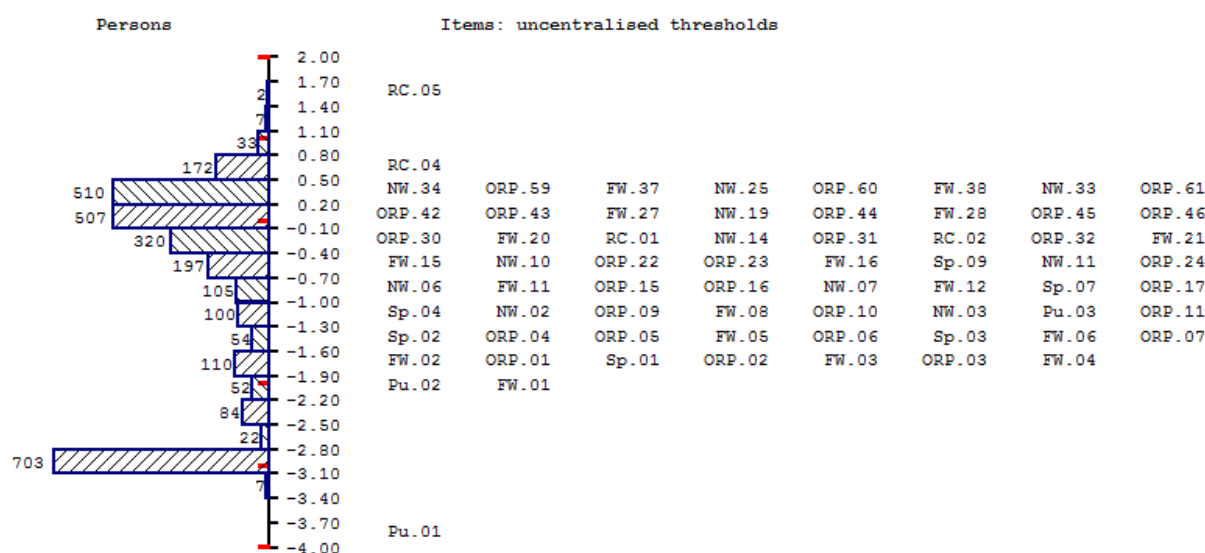
Setting the performance band boundaries

The next step was to add a short description of the skills required to answer each item on the full item location map. This helped to build up a picture of the competencies required to move up the scale. These appear to be broadly consistent with the order of competencies expressed in the standard one and standard two curriculum, although in many cases the curriculum statements are fairly general and are similar at the two levels. The difficulty location of the items (thresholds) for reading familiar words, and reading a short passage are in order of the number of words read per minute. Similarly the difficulty order is as expected for the reading comprehension, spelling and punctuating testlets (i.e. obtaining a higher score is more difficult). The exception is non-words where there is some threshold disorder at the higher reading speeds.

The demarcation of the five performance bands is in the table below. The location of the band boundaries is based on some key statements in the curriculum. First, reading comprehension is a standard two curriculum skill. As mentioned earlier, the government's BRN-education target for standard two pupils is reading at 50 words per minute. The location of the threshold of oral reading passage speed 50 marks the lower boundary of the top band 'achieving standard two level skills or above'. The location of the threshold of getting one reading comprehension question correct marks the lower boundary of next band down 'emerging standard two level skills'. The middle and second bottom band are termed 'achieving standard one level skills' and 'emerging standard one level skills'

respectively. It is the level of both reading and writing skills which mark the boundary between these two bands (see table for details). The lower boundary of the bottom band is set immediately below the location of the easiest item on the test (ignoring the first punctuation threshold[20]).

---

[20] As can be seen from the simplified item map above, the first punctuation threshold is very far below the other items. There are very few frequencies for this category and so it is assumed that the estimate for this threshold is not very reliable, and it is reasonable to ignore this.

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.
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. The spelling test included 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. The test included two fact-based questions.*
	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 impact evaluation team using data from IE baseline survey (pupil Kiswahili test)

I.3.2.2   Mathematics

Treatment of items
Each question on the mathematics test is treated as a dichotomous item which means there are two answer categories: correct or incorrect. Based on the Rasch diagnostic tests, one item was dropped from the number comparison subtest (ND6). Thus the number of items in each subtest is number comparison/ discrimination (7), missing numbers in sequences (8), addition (16), subtraction (16), multiplication (8), and word problems (4).

Treatment of non-response
All non-response is treated as missing.

Rasch analysis
The Rasch analysis of the mathematics pupil test produces estimates of person ability and item difficulty on an interval scale, summarised in the item location map below. The left-hand side of the scale shows the frequency distribution of the person ability estimates, while the right-hand side shows the location of the items.

The following abbreviations are used to label the items from each subtest: **ND**: number discrimination; **SEQ**: missing numbers in sequences; **ADD1**: addition level 1; **ADD2**: addition level 2; **SUB1**: subtraction level 1; **SUB2**: subtraction level 2; **MULT**: multiplication; **WP**: word problems
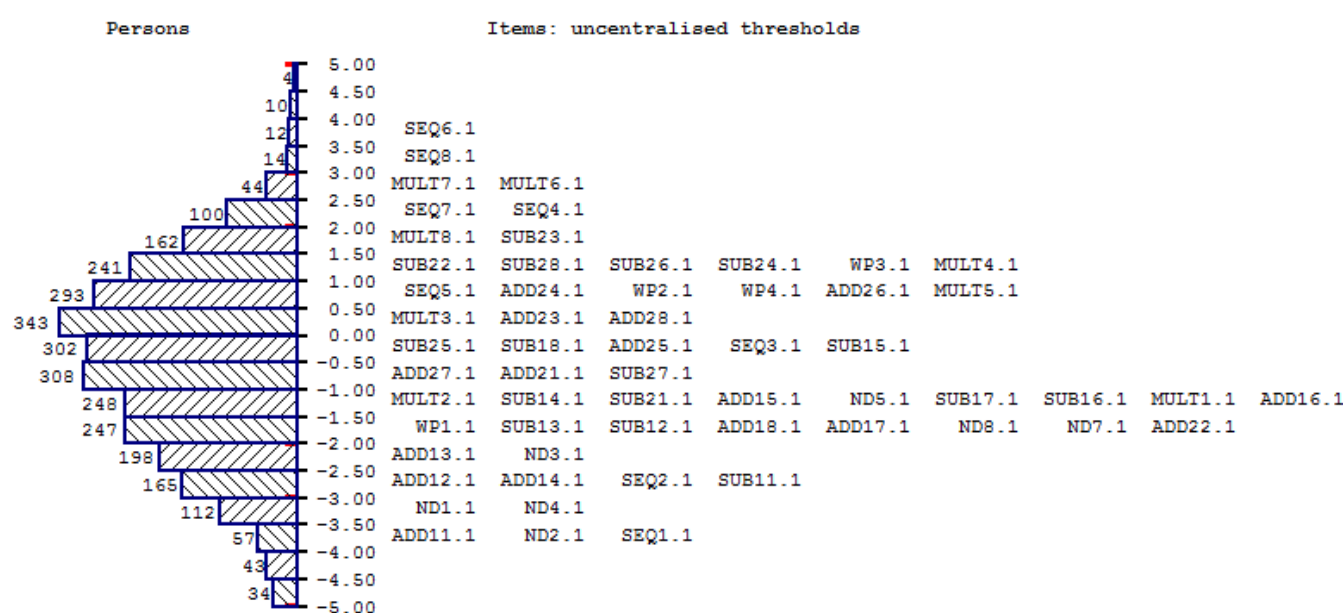
Process of setting the performance bands
The next step was to add a short description of the skills required to answer each item on the full item location map. This helped to build up a picture of the competencies required to move up the scale. These appear to be broadly consistent with the order of competencies expressed in the standard one and standard two curriculum.

The demarcation of the five performance bands is in the table below. The location of the band boundaries is based on some key statements in the curriculum.

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 down 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 impact evaluation team using data from IE baseline survey (pupil mathematics test). Notes (1): the items highlighted in small grey boxes (e.g. SUB1:7) are exceptions to the description given on the adjacent line.

I.3.2.3 Rasch model diagnostics

The sections below summarise the results from the diagnostics tests which were carried out on the pupil test data to assess its fit with the Rasch model.

SUMMARY OF RESULTS FROM THE DIAGNOSTICS OF THE PUPIL KISWAHILI TEST

Test score reliability: The person separation index (PSI, which is Rasch's equivalent of Cronbach's alpha) is high at 0.859 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 student ability estimate (unweighted mean = -0.912, standard deviation = 1.322). As the item map in the previous section shows pupil ability estimates were non-normally distributed, with a very large-floor effect, providing further evidence that the items were on average too difficult for the pupil cohort.

Item fit: This was examined using item characteristic curves (ICC). Before discussing the fit of the items in the final analysis, it is useful to explain why two subtests were deleted from the analysis. First, the listening comprehension subtest was deleted from the analysis, due to systematic misfit. The ICCs of the listening comprehension items (when treated as dichotomous items) showed very poor discrimination and a pattern similar to guessing on a multiple choice item (with two options given, the lower ability class intervals have around a .5 chance of getting the item correct). As an example, the ICC for listening comprehension question 4 is shown below, but this pattern was apparent across a number of the LC items, including LC2, 3 and 5. In fact, this result is unsurprising as when examining the substantive content of the item, the correct answer does not depend upon successful comprehension of the read passage, but rather on providing a logical response to the question which may be deduced from the comprehension of the passage, or may just be common-sense.
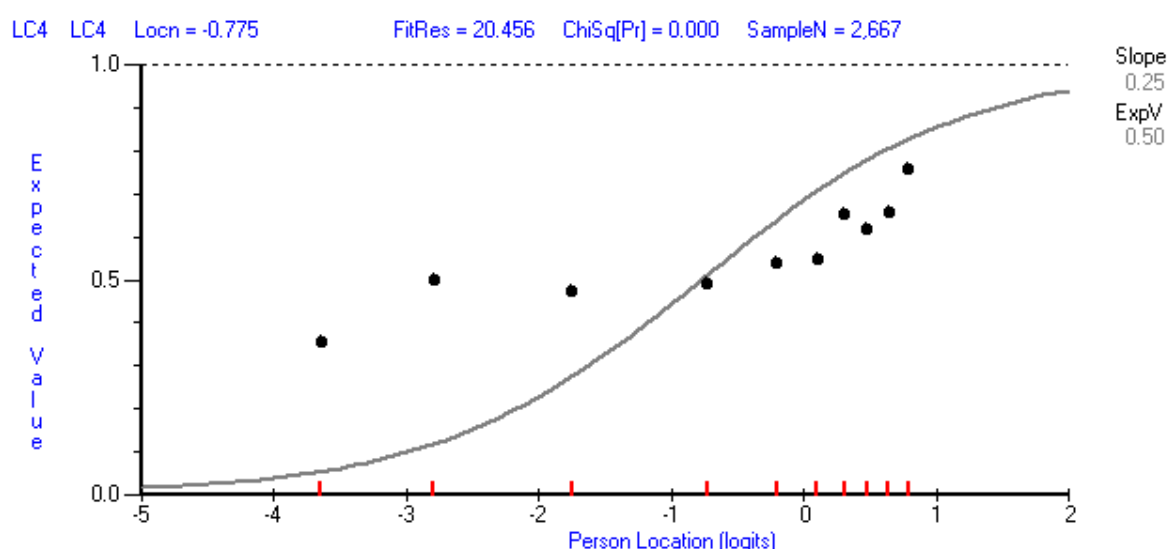
In addition, the syllables subtest was deleted as it was found to systematically discriminate less than the other items, it had disordered categories[21] and the principal components analysis of residuals provided evidence that this subtest was dimensionally divergent from the other subtests. These findings may be related to the way reading is taught in schools in Tanzania. If a phonic approach is not

---

[21] This means that the ordinal numbering of categories does not correspond with their substantive meaning. In this example, it meant that some slower syllable reading speeds were found higher on the scale than some faster syllable reading speeds.

at least part of the strategies employed by teachers to teach children to read, then the syllables subtest may be testing a different ability to early reading of whole words.

Turning to the results on the fit diagnostics for the final analysis. An examination of the item characteristic curves (ICC) showed that fit was generally reasonable. The best fitting item was the oral reading passage subtest, and its ICC is shown below.
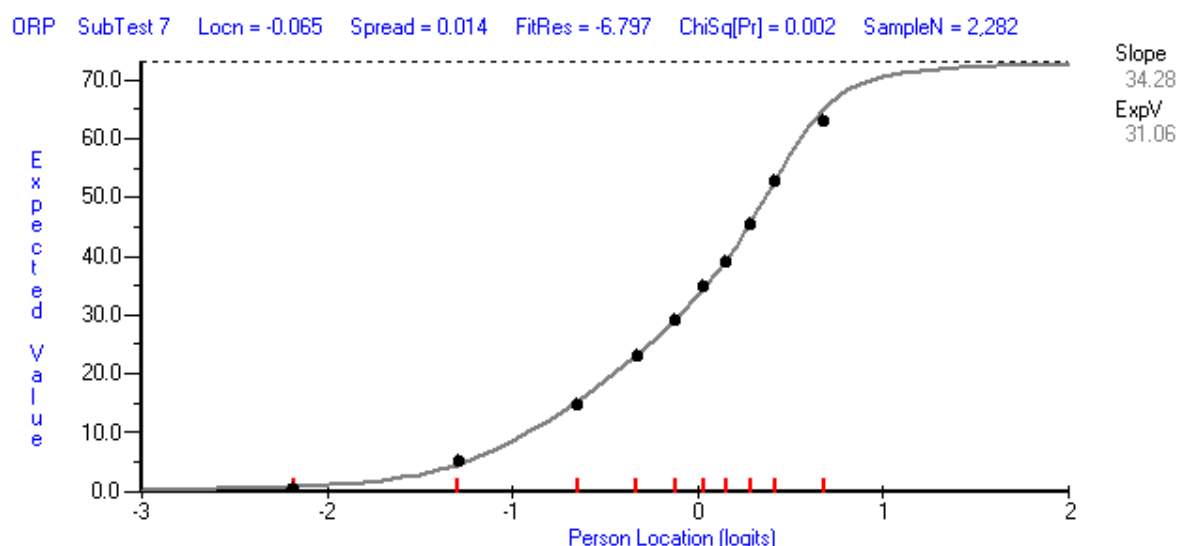
The worst fitting item was the reading Comprehension testlet. Its ICC is below.
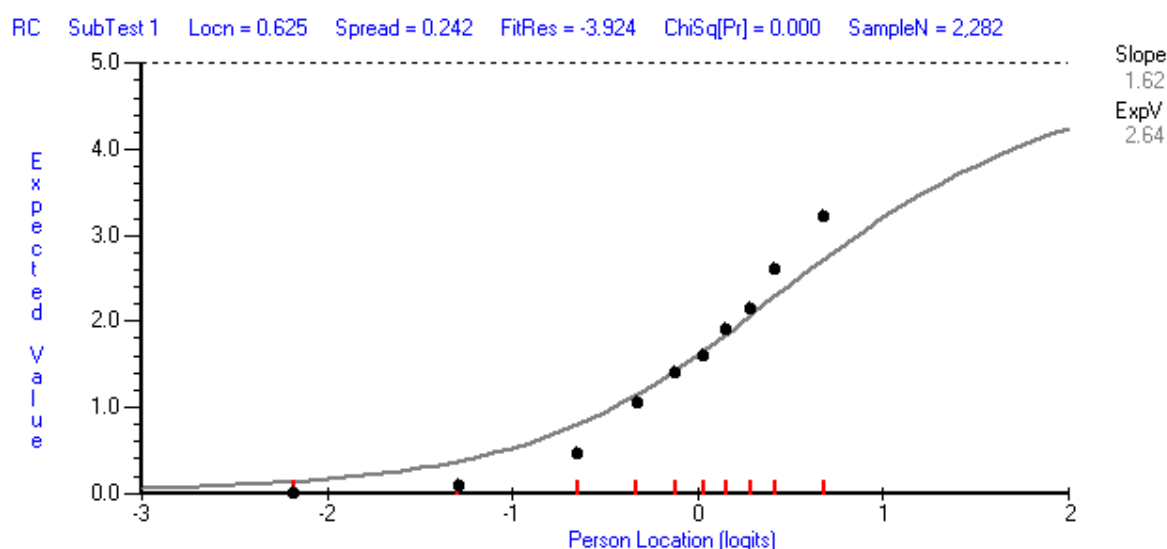
Clearly the reading comprehension item is discriminating higher than the average discrimination across all items. This is less problematic than poor discrimination and so the item was maintained, particularly as other tests of dimensionality did not flag the item as problematic

The category thresholds were disordered for the Non-Words subtest (see figure below) indicating that a higher score for this subtest did not necessarily correspond with higher levels of ability, as estimated across all items. However, this issue appears isolated to the very top categories where there are few people and so was not considered problematic enough to warrant deletion.

NW   SubTest 6   Locn = 0.030   Spread = 0.024   FitRes = 2.931   ChiSq[Pr] = 0.006   SampleN = 2,282
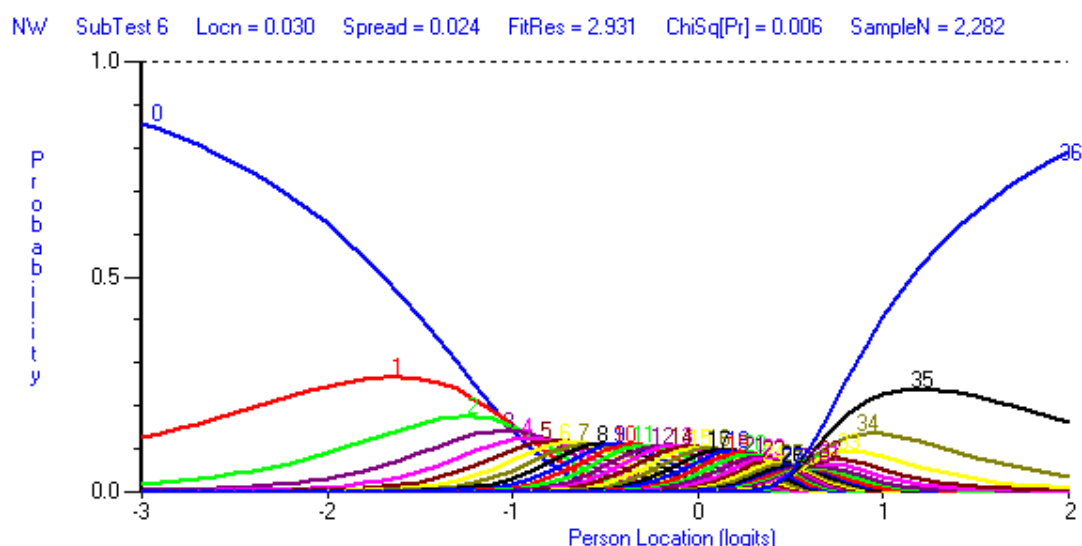
Item dependency: after combining the reading comprehension, spelling and punctuation items into three testlets, all residual correlations between items were acceptable. Furthermore, the principle component analysis of residuals did not provide any evidence of violations of unidimensionality.

Differential item functioning (DIF): All items were examined to see if they revealed DIF for four factors: gender (boys vs girls), treatment status (treatment vs control group pupils), home language (pupils who speak Kiswahili at home vs other local languages), and poverty status (poorer vs richer pupils). There was no significant DIF for the gender factor. For the treatment factor, the spelling testlet showed DIF such that pupils in the control group found this item systematically easier than pupils of equivalent ability (as estimated across all items) in the treatment group. For the reading comprehension testlet, there was evidence of DIF for both the treatment status factor and the poverty factor. Pupils who speak Kiswahili at home found this item somewhat easier than pupils of other local home language (of equivalent ability). Pupils from richer backgrounds found this item somewhat easier than pupils of poorer background (of equivalent ability). However, this DIF was less pronounced than for the home language factor and so this finding may just be reflective of the relation between wealth and home language. The substantive explanations for the DIF findings will be considered for future revisions of the test.

SUMMARY OF RESULTS FROM THE DIAGNOSTICS OF THE PUPIL MATHEMATICS TEST

Test score reliability: The person separation index (PSI, which is Rasch's equivalent of Cronbach's alpha) is high at 0.93 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 student ability estimate (unweighted mean = -0.541, standard deviation = 1.809). As the item map in the previous section shows pupil ability estimates were quite normally distributed, but there is a very slight skew towards lower ability levels.

Item fit: This was examined using item characteristic curves (ICC). Before discussing the fit of the items in the final analysis, it is useful to explain why one item was deleted from the analysis. The ICC for item 6 on the number discrimination subtest (see below) shows clear guessing, and poor, and at times negative, discrimination (the only item on the test to do so).
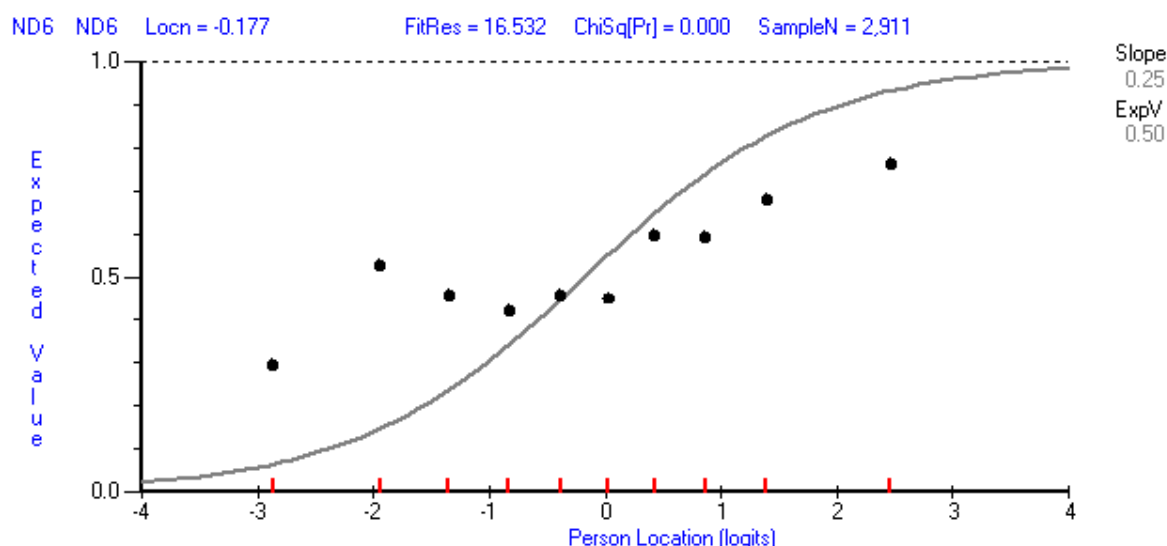
For the final analysis, generally the ICCs show a reasonably good fit to the Rasch model. The best fitting item on the test is word problem 4, and its ICC is below.
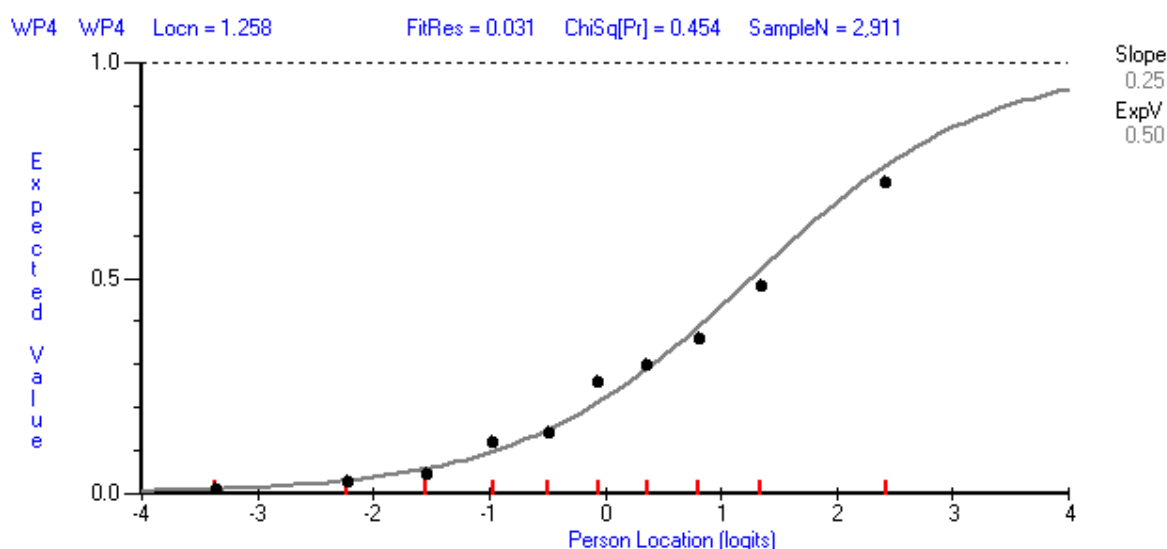
Item dependency: Overall the residual correlations overall look acceptable. There is some evidence of dependence between Seq1 and 2, Seq6 and 8, Add1-1 and 2, the Add2 subtest, Sub1 subtest, Sub2 subtest and Multi subtest. This is unsurprising given the similarity between items within the subtests. No patterns appear severe enough to warrant dropping or combining items/subtests.

Differential item functioning (DIF): There was no substantial DIF for the gender factor or for the treatment status factor.

Annex J   Details of qualitative sampling, tools, fieldwork and data coding

J.1   Selection of research sites

J.1.1   Stage 1: Selection of districts

The quantitative survey sampling frame was also used to select districts for the qualitative research. Control group districts were removed as possible sites as the qualitative research is primarily concerned with the changes within programme districts.

Three districts were selected using purposive 'typical case sampling' and 'extreme case sampling' on education inputs, outputs and outcomes at the district level, with consideration given to the social and economic resources of the population in each district.

The strength of relationships between education outcomes and key social and economic pupil background characteristics was analysed (Pearson correlation coefficient), using UWEZO and national examinations data. The variables with the strongest relationships with performance were used to develop a pupil social and economic resources indicator. The strength of the correlation was interpreted using Dancey and Reidy's (2004) categorisation. These indicators were selected on the basis of the relative strength of relationship, as compared with other indicators. Economic indicators included having electricity at home and ownership of a fridge. Social resources included mother having a secondary education and father having a secondary education. The number of books in the household was also included on the basis of the international research regarding the importance of books in the home over the lifecycle. Each indicator was given equal weighting for the composite measure.

The strength of relationships between education outcomes and key school background characteristics was analysed. The variables with the strongest relationships with performance were used to develop a school resources indicator. School resources indicators included the pupil-teacher ratio and per capita spending per primary school pupil. Each indicator was given equal weighting in the composite indicator.

The performance, school resources and pupil social and economic indicators for each district were then mapped into four categories, according to their rank against all districts. Those districts within the lowest quartile of districts were marked red, those districts with a value within the third quartile of all districts were marked orange, those districts with a value within the second quartile of all districts were marked yellow and the districts in the top quartile were marked green.

The districts identified as being at risk of contamination from other programmes were then removed as possible case study sites.

Each district was 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'. This is based on the assumption that districts performing well, typically and worse than expected, with diverse pupil and school level resourcing, will enable the research to capture heterogeneous impact.

The final district selection was based on the following criteria:

- One district from each of the above categories must be represented, in line with the qualitative research sampling design, requiring typical and extreme case sampling (high, typical and low performing districts);

- Pupil performance (using UZEWO scores) must represent three of the four performance quartiles at the district level, in order to ensure that the typical and extreme cases cover absolute performance, not only expected performance based on resources;

- The level of pupil and school social and economic resourcing should be diverse in order to provide the highest chances of heterogeneous impact;

- The three districts must be taken from different Regions in order to provide as much coverage of diver context across the treatment districts as possible; and

- Travel time between the three locations should be practical and represent Value for Money, given the time and resources constraints on the evaluation.

Given the importance of gender to EQUIP-T, the selected districts were then reviewed in light of lower secondary enrolment by gender. Two of the selected districts had higher than average female enrolment in lower secondary education and one had significantly lower enrolment. The three selected districts are shown in Annex table 28.

Annex table 28 Selected districts

Region	District	Literacy (Kiswahili) Std. 3	School Level	Social and economic resources	Performance category	% female enrolment in lower secondary education
Dodoma	Mpwapwa DC	13	93	0.31	Performing worse than expected	41.9
Tabora	Uyui DC	17	9	0.46	Performing better than expected	42.9
Shinyanga	Kishapu DC	21	94	0.37	Performing typically	34.8

J.1.2 Stage 2: Selection of schools

Three schools within each of the three selected districts were selected using purposive 'typical case sampling' and 'extreme case sampling'. Control group district schools were removed as possible sites.

PSLE 2012 school ranks were downloaded from the National Examinations Council (NECTA) website and on the basis of each school's 2012 PSLE rank, each programme school was given a rank from 1-8, against the other schools in the applicable district. Kishapu ranks were not available, so this was only done for Mpwapwa and Uyui.

The 2010 average GPA for Maarita and Kiswahili was then subtracted from the average 2012 GPA for these subjects for each school in order to provide an indicative academic improvement variable to ensure the schools did not only represent diversity in relation to relative performance (as compared to other schools in the district), but also in relation to the rate of improvement. The measure is indicative as it is open to cohort effects because this measure only takes two years of results into consideration. However, an indicative improvement variable is important as the more

diverse the school case studies, the more likely it is that any heterogeneous impact of the EQUIP-T programme will be identified.

A pupil-classroom ratio variable was calculated by dividing the total grade 1-7 enrolments by the total number of classrooms and a percentage of teachers qualified variable was calculated by dividing the total number of qualified teachers by the total number of teachers. However, each school was found to have 100 percent qualified teachers, based on the EMIS data. A pupil-teacher ratio variable was computed by dividing the total grade 1-7 enrolments by the total number of teachers.

Each school within each district was then ranked from 1-8 on the basis of the pupil-teacher ratios and pupil-classroom ratios.

Annex table 29 School case study sites in Mpwapwa, Uyui and Kishapu

District	School	Performance	Academic improvement	Gender Parity
Mpwapwa	Mpwapwa School A	Performing typically	Medium to high	Not close to gender parity
Mpwapwa	Mpwapwa School B	High performing	High	Almost achieving gender parity
Mpwapwa	Mpwapwa School C	Not performing well	Low	Not close to Gender parity
Uyui	Uyui School A	Typical performance	Above average	Not close to gender parity
Uyui	Uyui School B	Not performing well	Low	Closer to gender parity
Uyui	Uyui School C	High performing	High	Closer to gender parity
Kishapu	Kishapu School A	Typical performance	Low	N/A
Kishapu	Kishapu School B	Good Performance	High	N/A
Kishapu	Kishapu School C	Not performing well	High	N/A

J.2 Description of focus group discussion and key informant interview tools

The FGD with parents covered a broad range of topics, including the benefits of education, challenges with education in their locality and within their children's school, the characteristics of a good school, and the characteristics of a good teacher. Parents were asked to discuss what can be done to improve schools and teachers and who is responsible for doing so. Parents discussed how to ensure that children learn, and the roles and responsibilities of various stakeholder groups in ensuring that children learn. Two participatory tools were at the disposal of the team and could be used if needed for the discussion. Vignettes were used to explore what happens when parents come to school to complain about a specific topic, and when a teacher arrives at a school without speaking the local language. A problem tree analysis was used to elicit the views of parents on the problems of education for children in standard III and the possible causes and consequences for the education of girls and boys. These exercises were particularly useful in communities where Kiswahili language skills were low.

FGDs with teachers explored a range of issues, including the reasons participants became teachers, the aspirations and expectations of teachers when becoming teachers, perceptions of the teaching profession, the responsibilities of different actors in the functioning of school life, the qualities necessary to be a good teacher, challenges teachers face and possible solutions for overcoming these challenges. Scenario discussions (vignettes) were used to discuss questions around girls pass rates in standard VII, deployment of teachers to rural areas without speaking the local language and the classroom absenteeism of teachers.

The FGD with the SMC focused on questions around the role and responsibilities of the SMC, the division of these responsibilities, the challenges of fulfilling these responsibilities, possible ways of

addressing these challenges, the division of responsibilities within the school and between the school and the community, the qualities of a good teacher and possible ways of supporting

---

Participatory research methods used with children

A day in 'my town' The aim of this exercise was to elicit the views of children on how boys and girls differ in terms of experiences of school and life, if any. This was done by asking children to describe a normal day that girls/boys of their age experience and how that experience differs and why. 'Growing up in [name of town/village]' aims to elicit information on the expectations of girls and boys within the community, including what they are expected to do, what they expected to become and the role of education and schooling. The activity generated a shared timeline of the major events and experiences of a boy and a girl within each case study location.

Scenario or vignette discussion This exercise sought to understand children's perspectives on the quality of their education and experiences of schooling. Vignettes were used to elicit children's views and experiences at school. The facilitator read an incomplete story and the group completed it orally, with the facilitator prompting the children to continue filling in the storyline until the story came to a comfortable conclusion. Topics covered depended on the locality, but included retention of girls in the latter years of primary schooling, attitudes towards female teachers, school absenteeism, teachers who only speak Kiswahili and taking a letter from school home to parents.

Forcefield analysis of school Good and bad things about schooling and school: the aim of this exercise was to generate information on children's views of school using two forcefield exercises. The aim is to elicit children's thoughts on opposing forces in relation to their school: good forces, represented by helium balloons, pulling the school upwards, and bad forces, represented by stones, pulling the school downwards. The school remains on a line in the middle of a flip chart. We were interested in children's views of both their specific school and education in general.

---

teachers to become better teachers. The FGD also included an exploration of the possible mechanisms available for ensuring that quality learning takes place. Scenario discussions around similar questions as in the teacher FGD could be used if necessary.

J.3   Fieldwork, data coding and analysis

J.3.1   Selection of national researchers

The qualitative team members were selected on the basis of having relevant research experience or knowledge of the education system in Tanzania. Initially the qualitative team had sought to work together with members from the University of Dar es Salaam's faculty of education, but the change to the timing of the fieldwork meant that this initial team was no longer available for the research.

The local team was extremely hard working and dedicated and showed a great interest in the project, going well beyond the data collection. Despite the plethora of research activities that are currently taking place in Tanzania, each organization has its own standards for what constitutes quality research and the team had to be trained to meet OPM standards. Whilst English is widely spoken and the training was conducted in English, the risks associated with the possibilities of losing some of the nuances of the training was addressed by making extra time during training and debriefs in order to ensure that all issues were adequately discussed and understood.

J.3.2   Daily debriefs and team checks in the field

As a key part of the process, teams conducted some initial synthesis and analysis in the field. This started at the level of the FGDs or interviews, but also occurred at the school and community, as well as the district level. The aim was to conduct thorough debriefs and initial analyses of findings to feed into the draft baseline report, further refine the research tools, discuss interesting emerging issues for further exploration and jointly conduct an initial analysis of the data.

The team was given a debrief format and asked to prepare both a description of the school and the environment, as well as debrief notes at the end of each day. Each team was then asked to make sure that the recordings were safely copied and that digital photos of the visual outputs from the participatory tools existed. In the daily debriefs, each team would present on their findings of the day which were then discussed and recorded by the group. Research gaps that needed addressing during the next day were identified and any possible biases discussed on a daily basis.

J.3.3 Data treatment and collection

All team members were asked to write fieldwork journals which were used to feed into the daily debriefs. Debrief reports exist for each school, as well as brief descriptions of the schools and the environment surrounding it. The fieldwork plan was constantly updated to allow for more accurate information for the next wave of the qualitative research. In addition, the team leader and OPM staff member accompanying the fieldwork also kept a fieldwork journal.

All FGDs were recorded, as well as most key informant interviews. Recordings, visual outputs and consent forms are safely stored for further analysis.

J.3.4 Coding for analysis

After the data was collected and transcribed, the team built and applied codes by indicator, TOC mechanism and contextual information. Data from KIIs and FGDs were transcribed, translated (if required) and coded.

An excel spreadsheet was developed with one tab per qualitative indicator across the four programme components. These indicators included: perceptions of learning, communication between school and home, attitudes to female teachers, perceived reasons for learning differences between groups, attitudes to schooling boys and girls, teacher morale, capacity and behaviour, teacher attitudes to girl pupils, school leadership performance, school leadership attitudes to girls, effectiveness of district planning, understanding of roles and responsibilities, engagement with the community, implementation information, contextual information. New tabs were added as areas of interest from the data emerged. For example, Perceptions of 'school readiness'.

A freeze window was created for every qualitative indicator tab including descriptions and corresponding codes for level of participant group, participant group, region and district and school.

Annex table 30 Descriptions and codes for each qualitative indicator

Levels		Groups		Region/district		School	
National	1	EQUIP Staff	1			Mpwapwa School A	1
Regional	2	REO	2	Dodoma/Mpwapwa	1	Mpwapwa School B	2
District	3	DEO	3			Mpwapwa School C	3
Ward	4	WEC	4			Uyui School A	4
School	5	HT	5	Tabora/Uyui	2	Uyui School B	5
		SMC	6			Uyui School C	6
		Teachers	7	Shinyanga/Kishapu	3	Kishapy School A	7

	Pupils	8			Kishapy School B	8
	Community Leaders	9			Kishapy School C	9
	Parent	10				
	Academics	11				

Below the descriptions and codes the name of the indicator was listed, followed by several columns indicating where codes and data from the transcripts should be populated. As researchers populated the cells with perception data across all of the indicator tabs from each transcript, the level, participant group, region/district, school and any pertinent comments were added to the row. Once several stakeholder groups were added to the analysis sheet, researchers began identifying key responses across the perception data and adding creating response codes to be used across the indicator.

Annex table 31 provides a fictional example of a row for the indicator *'communication between home and school'*.

Annex table 31 Fictional example of codes and data population cells

Communication between home and school						
Response code	Level	Participant group	Region/district	School	Perception data	Comment
A	4	4	2	4	There is no school notice board in that school. In other schools the notice board is in the HT's office so very few people will have access. Only me, as the WEC.	

Response category	Response Code
Notice boards	A
Discussions between parents and teachers	B
School/community meetings	C
Through the community leader	D

In addition to the qualitative indicator tabs, five further excel tabs were created corresponding to the TOC for each of the four programme components, and including one unstructured TOC tab for cross-component data or previously not considered aspects of the TOC. These five tabs also included descriptions and corresponding codes for level of participant group, participant group, region and district and school, in exactly the same way as the qualitative indicator tab freeze panes.

Below the descriptions and codes the name of the programme component was listed, followed by several columns indicating where codes and data from the transcripts should be populated. As researchers populated the cells with data pertaining to the TOC for each programme component, the level, participant group, region/district, school and any pertinent comments were added to the row. As data were added to the analysis sheet, researchers coded each response by assumption, as identified in the qualitative evaluation matrix. Annex table 32 provides an example of response codes by assumption for the community component of the EQUIPT programme.

Annex table 32 Assumption and response codes for the community component qualitative data analysis

Assumption	Response Code

Community sensitisation and a community education plan will increase buy-in by parents which will in turn ensure that the community is more supportive of schools	1
SC currently do not have the necessary skills to fulfil their intended functions (PFM, planning, management and knowledge of their statutory role)	2
Communities rarely hold school management and teachers to account for the delivery of education	3
An informed community will demand that SC are representative of it and reflect its wishes in the school management and planning process	4
An active community will ensure that pressure is applied on the SC and school to report to the community on their actions. This demand for accountability will ultimately be passed up the chain, leading to increased reporting and accountability throughout the entire chain of education management from the district level down.	5
The development of an education plan by communities will ensure that needs and wishes identified by the community are taken up by the SC and reflected in the school management plan. School management plans will thus be demand driven and reflect community priorities	6
A SC that is representative of the community and has received training to fulfil its intended functions will ensure that the education plan developed by the community is reflected in the school development plan	7
Other	8

These data could then be sorted in the excel spreadsheet. Sorting was undertaken against each qualitative indicator and TOC for each component by stakeholder group, case and group. The team then applied a confirmatory analysis and/or description of phenomenon within individual cases (schools), compare individual cases (schools), compare groups of cases (wards) and test and validate the TOC across cases (Based on Ryan's 'Standards of Rigor for Qualitative Research').

Annex K   Qualitative evaluation matrix

Annex table 33 Qualitative evaluation matrix

Area of impact	Hypotheses	Assumptions (output to outcome level)	Core areas to probe (including change over time for each)	Source of information	Qualitative indicator
Output 1: Enhanced professional capacity and performance of teachers	INSET teacher training and a management performance system linked to career development lead to improved teaching	• Teachers, head teachers and/or INSET coordinators will through the EQUIP-T mixture of outputs be more willing and able to take on new skills, roles or ways of working • A monitoring system that rewards performance and links to career development will improve teaching quality • High quality training can improve teaching quality • Mainstreaming gender in INSET training will improve gender equity	• How the INSET teacher training is implemented in the region • Perceived benefits and challenges of working as a teacher • How existing challenges affect teachers attitudes and behaviour at school • Factors that would contribute to doing a better job as a teacher (probe importance of training in relation to other factors) • Satisfaction with acquired skills and skills that can be improved (including gender sensitive pedagogy) • Willingness and ability to use new skills from received training (including gender sensitive pedagogy) • Existing procedures in place to motivate teachers and how useful these are for improving teaching quality • Perceived role and usefulness of teacher appraisals for improved teacher performance (OPRAS and the new TPF)	• KIIs with head teacher and designated INSET teachers (ML and EL only) • KII with community leader • FGDs with teachers • FGD with parents • FGD with SMC • FGDs with pupils • KIIs with key district and ward officials • EQUIP-T programme managers	• Perceptions of teacher capacity and performance • Perceived explanations of change in teacher capacity and performance • Teacher attitudes to female pupils
Output 2: Enhanced school leadership and management skills	Leadership performance will be improved by helping leaders know what is expected of them, setting targets for performance, and offering professional development to help them achieve those targets.	• School leaders find the attended training appropriate/fit for purpose. • School leaders will apply what they have learnt to improve their own performance and the wider performance of the school. • The DEO and school leaders will want the Peer Support Network to continue after EQUIP-Tanzania Leadership Professional Development has completed. • The District Education Offices are	• How leadership and management is implemented in the region • How school leaders understand their roles and responsibilities • Areas of satisfaction and frustration experienced by school leaders • How these frustrations affect school leaders attitudes and behaviour at school • How existing challenges can be addressed (probe relevance of, commitment to and effectiveness of EQUIP-T outputs) • Perceived role and impact of appraisals to	• KII with head teacher • KII with community leader • FGDs with teachers • FGD with SMC • KIIs with key district and ward officials	• Perceptions of school leadership performance • Perceived explanations of change in leadership performance • School leaders attitudes to female pupils

		committed to the Performance Management System	improve the performance of school leaders (including for improved gender and equity) • Perceived effect of management and leadership training for overcoming the challenges school leaders face (were the training fit for purpose, to what extent has the training been applied, in what ways and with what effect)		
Output 3: Strengthened systems that support the district and regional management of education	Capacity building at the district and regional level will lead to strengthened systems and improved education outcomes	• District Offices release grants received in full and on time and this will give head teachers more certainty and make them lead school planning more decisively • Timely submission of school-based data to districts will make it possible to use EMIS data for planning purposes • The development of a district education plan will contribute to better district education planning • Personnel who have acquired new skills are willing to change behaviours and put their learning into practice • Political elites and vested interests will not disrupt or block elements of the programme • EQUIP-T outputs will make district officials more willing and able to deal with sensitive issues (e.g. abuse of girl pupils/ women teachers, early pregnancy) as part of the district education planning cycle	• How regional and district management support is implemented in various regions and districts • How district and ward officials understand their roles and responsibilities • Challenges district and ward officials face in fulfilling their responsibilities • Effects for schools of existing challenges • Perceived strengths and weaknesses with the district education planning cycle (including planning, budgeting, execution, monitoring) • Perceived frustrations with the management of the capitation grant • Perceived ability to use EMIS data for planning purposes • Factors that would improve the functioning of the district planning cycle • Perceived effect of capacity building activities for improving the planning cycle and releasing capitation grants in full and on time	• KIIs with key district officials • KIIs with Ward Education Coordinators • KII with head teachers • KII with community leader • EQUIP-T programme managers	• Perception of effectiveness of district education planning • Perceived explanations to changes in the district education planning • Availability of costed plan for the long term maintenance of quality within schools
Output 4: Strengthened community participation and demand for	Greater engagement by the community in school operations and outcomes will provide much needed support and resources, that will	• The use of school performance monitoring tools will make communities adequately informed about school operations, needs, plans and performance • Training will make community	• How community participation and accountability in education is implemented in each region/district • Understandings of the roles and responsibilities of the school committee • What the school committee is good at and where it can improve (including	• FGDs with parents • FGDs with pupils • FGD with SMC • FGD with teachers • KII with head teacher	• Perception of engagement by the community in the school • Perceived

| accountability in education | enhance the relevance of education | stakeholders aware of their rights and responsibilities
• Training of the school committee will improve the capacity to plan and manage financial resources
• Financial resources for education will be generated by upgrading existing school based IGAs or establishing new IGAs. | whether represents community views and wishes)
• Opinions about the usefulness and impact of training of school committee members
• Community/parents views on the access to information about the school and its performance (adequacy, benefits)
• Perceived ability of the community/parents to hold school management and teachers to account for the delivery of education
• Perceived role and usefulness of the whole school development planning
• Views on the relevance and utility of the PTA
• Attitudes towards supporting the school (financially, in-kind, other types of support)
• Perceived usefulness and transparency of income generating activities
• How perceived weaknesses in community participation can be overcome (probe relevance of EQUIP-T activities for improved participation and accountability) | • KII with community leader
• EQUIP-T Programme managers | explanations to changes in community engagement in the school
• Attitudes to female pupils |

Annex L   Key qualitative research questions and data sources

Indicator	Baseline	Midline	Endline	EQUIP Staff KII	REO KII	DEO KII	WEC KII	HT KII	SMC FGDs	Teachers FGD	INSET Teacher FGDs	Pupils FGDs	Community Leader KII	Parents FGDs
From Evaluation Matrix/Analysis Plan														
What are perceptions of overall learning levels, and any differences between groups of pupils?	X	X	X	X	X	X	X	X	X	X		X	X	X
What are perceived reasons for overall learning levels	X	X	X	X	X	X	X	X	X	X		X	X	X
What are attitudes to schooling for boys and for girls?	X	X	X											
Perceptions of reasons for differential pupil performance for boys and girls	X	X	X	X	X	X	X	X	X			X	X	X
Perceptions of reasons for (any) changes in performance		X	X	X	X	X	X	X	X	X	X	X	X	
Perceptions of teacher morale, capacity and performance	X	X	X	X	X	X	X	X	X	X	X	X		
Perceived explanations for change in teacher capacity and performance		X	X	X	X	X	X	X	X	X	X	X	X	X
Teacher attitudes to female pupils	X	X	X				X	X	X	X	X	X	X	X
Perceptions of school leadership performance	X	X	X				X	X	X	X			X	X
Perceived explanations of changes to leadership performance							X	X	X	X			X	X
School leaders' attitudes to female pupils	X	X	X				X	X	X	X		X	X	X
Perceptions of effectiveness of district education planning	X	X	X	X	X	X	X	X	X				X	
Perceived explanations to changes in district education planning		X	X	X	X	X	X	X	X				X	
Perceptions of engagement by the school community in the school	X	X	X				X	X	X	X		X	X	X
Explanations of changes in community engagement in the school							X	X	X	X		X	X	X
Attitudes to female teachers	X	X	X				X	X	X	X		X	X	X
Perceived changes in attitudes to female teachers							X	X	X	X		X	X	X
What are perceptions of communication between the school and community?	X	X	X				X	X	X	X		X	X	X

Annex M   Qualitative analysis background paper

M.1   Introduction

The qualitative research design and the rationale for selecting districts and schools for this work have been described earlier in this volume (Chapter 4 and Annex J). This Annex provides a more detailed narrative of each of the three districts chosen for the qualitative research and the three schools chosen in each district under the headings: teacher capacity, morale and behaviour; school leadership and management; and relationship with the community.

The aim of this section is to provide a case by case description of the research sites that augments the across cases analysis presented in Volume I.

M.2   Uyui district, Tabora

Uyui, Tabora was selected as one of three qualitative case study districts on the basis of the available quantitative data on literacy, school resources, social and economic resources and the GPI at the lower secondary level (See sampling strategy in Annex J). Uyui is a district performing slightly better than other districts, despite low levels of school resources and typical levels of social and economic resources. The percentage of girls enrolled at lower secondary level in Uyui is reasonably high compared with other EQUIP-T intervention districts. Uyui is one of the districts in the Tabora region.

Three schools were purposively selected in the Uyui district. Among the nine schools, those in Uyui are relatively large with large classes and high pupil-teacher ratios.

M.2.1 Uyui Better Performing Primary School (Uyui Better)

Uyui Better is a medium-sized school (454 pupils) with fairly large class sizes and pupil-teacher ratios (both 65). The standard one to three pupil absence rate is about average (36%) and is higher for girls than for boys. The head is female and has been in the post for three years. Uyui Better is located in the Lolangulu ward and was selected as a relatively well performing school with high, recent improvement. The school is closer to gender parity than other schools in the district.

The community suffers from inadequate access to water, with the village wells drying up. Children are often asked to bring water to school for cooking, cleaning of toilets and gardening, whilst tapped water for drinking purposes is available in the village. The community are agriculturalists and cattle keepers. Despite income generating activities (IGAs), the school is typically unable to provide school feeding for more than a couple of months per year.

M.2.1.1   Teacher capacity, morale and behaviour

Teachers at Uyui Better described fond memories of joining the teaching profession and gaining a qualification, despite most of the teachers stating that they joined the profession because they had no alternative. However, their experiences of the profession did not match their expectations in a range of areas. These areas included no career progression, low and late salaries, no training, administrative errors at the district level, a perceived lack of respect from the community and government, too many responsibilities, a lack of essential teaching and learning materials, no teachers housing and the frequently changing curriculum.

As with all of the nine case study schools, a lack of teacher housing was perceived to lead to more interruptions to instructional time than any other factor. Teachers claimed that they are late to school

because they need to walk long distances, often missing the first three periods. However, the head teacher reported that if allowances are made for teachers to arrive late, due to this constraint, teachers arrive after noon. The management of teacher's arrival times was a time-consuming responsibility for the head teacher. The community found previous teachers to be lazy and judged that they did not 'deserve' housing. However, the community perceives the current teachers as deserving and expressed a desire for the teachers to live within their community with the exception of one teacher who is frequently drunk whilst at school. The community has been unable to have him removed or hold him to account despite attempts to do so. Nonetheless, parents would like the teachers to live in the community as they thought that this would allow them a greater level of control over the lives of teachers and hold them to account more readily for adverse behaviour. The community has developed a plan to build houses for the teachers, but this plan has been discussed for many years without being implemented. The community claims that this is due to the fact that they cannot manage this task without government support. The situation is currently at an impasse and causes tensions between teachers and the community.

The capacity of teachers to perform their duties was constrained by the changing curriculum and in particular, the introduction of topics that were perceived not to be useful for rural areas; for example Information, Communication and Technology. One of the most common complaints by teachers included the time required for preparing classes. While this may be in part due to the large number of periods to be covered, this may also indicate lower levels of capacity in pedagogy, classroom management and subject knowledge. Understaffing seems to be a problem at this school which leads to teachers being overstretched and having too many classes to teach. There are also reports of teachers having overlapping schedules. However, the head teacher and teachers did not agree with this viewpoint. According to the head teacher there are only six classrooms and seven educators. As a result she claims to know at every point in time where each teacher is and ought to be. In addition, late, incomplete or no salary payment was reported to result in teachers engaging in other activities to supplement their income, leading to teacher absenteeism. The adage '*you have failed even to become a teacher'*, describes a common perspective regarding the standard of the profession. This was cited by education officials in multiple districts and by teachers within Uyui Better. Furthermore, teachers also consider some tasks they are responsible and accountable for such as reporting thefts at school or tending to injured children as tedious and time consuming.

> "I hate this job especially when I am on duty … taking care of the children and school property. That means that I am getting tired and that I do a lot of work at a time. For example, if a theft has occurred then I will be accountable to state what happened."

> "If a teacher found that the child got an accident/trouble at the school compound, he/she has to stop teaching and assist the child. Or if a theft has occurred then you will be responsible so you spend a lot of time on solving problems."

The lack of training provided since joining the profession contributed to both low levels of capacity and morale. Teachers also know that teaching methods are changing and that they are being left behind, without adequate training being provided. Teachers expressed a general sense of feeling abandoned by the government. When a teacher was granted permission to undertake training, administrative errors at the district office resulted in many more weeks' loss of instructional time when the approval went missing and the teacher was falsely reported as absent from school and suspended.

M.2.1.2   School Leadership and Management

During a facilitated scenario discussion teachers reported that head teachers should respond to a teacher arriving late at school by helping the teacher solve his or her problem. However, the head teacher stated that if she is not strict with teachers, they will arrive even later. The SC tries to incentivise the teachers by writing down achievements, however they do not have the resources to

provide stronger incentives. While the head teacher discusses being strict, both the teachers and the head teacher at Uyui Better viewed the authority of the head teacher only in terms of the ability to report up to the school inspector. In addition, teachers discussed differential behaviours for school inspection, in that teachers perceived different behaviours as acceptable when no inspector is present.

The WEC was reported to have visited twice in six months, but was in regular contact with the head teacher whom he met in town. The community, on the other hand, did not think that the WEC was responsive to their complaints. A lack of transport was mentioned as restricting the WEC in his ability to do his job, leading him to summon teachers to his office for meetings rather than visiting the school himself. The community leader felt that the WEC should visit the school at least once a month and not every three months as is currently the case. He felt that only if the WEC comes to school is he able to really assess the challenges faced on a daily basis.

Uyui Better is currently experiencing a power struggle between the former and current head. The former head teacher went for further studies and was replaced by the current head teacher. However, he has now returned to the school as an ordinary teacher which leads to conflicts as he attempts to get his job back and both the current and the former head teacher are trying to run the school. This matter has already been raised with the district and the WEC was called in and notified of the problem. The community suggested that the former head teacher should be transferred to another school in order to solve the problem. The community views the gender of the new head teacher as beneficial for fulfilling her role, but thinks that her gender is partly to blame for the behaviour of the former, male, head teacher towards her. In addition, the fact that the qualifications of the former head teacher are higher than those of the current head teacher was perceived as further adding strain to their relationship.

Whilst the community leader feels that he can call meetings with the WEC and DEO, he also thinks that they do not listen to him and the community, as is evidenced by the lack of action around the conflict between the old and the new head teacher. Nonetheless, he conceptualises his role as being responsible for reporting teacher absenteeism to the WEC and to the district level.

M.2.1.3 Relationship with the Community

Besides housing for teachers, the biggest causes of conflict between the community and the school were the discipline of children, the quality of education and the behaviour of teachers.

Teachers discussed hitting standard one pupils on the palms with sticks and standard six pupils on the buttocks, but claimed that the procedures (limiting punishments to three sticks) outlined by the government have been designed by officers who are unaware of life as a teacher in the real world. The teachers, therefore, agreed on a new procedure whereby very thick sticks can be used, which would be more comparable to a ten sticks beating with the typically used, thinner sticks. Parents were not convinced that their children deserved this type of punishment, yet felt that if they requested a rationale for 'beatings' teachers will deny their child attention in class. Some parents reported coming into school and demanding to know why their children were being punished, whilst others felt that corporal punishment was acceptable, as long as a clear reason was given and the fault was the child's. Nonetheless, the issue of corporal punishment seems to occupy both parents' and teachers' time and leads to conflict between the school and community. Another source of conflict is teacher absenteeism and the quality of teaching which parents find lacking. Teachers on the other hand are frustrated by the lack of support from the community in identifying the men involved in getting female pupils pregnant. The community shields these mostly elder men, whilst teachers want to hold them to account. A lack of reporting on financial income and expenditure by the school leads to suspicion, especially around IGAs, as parents want to know whether the money made benefits their children or is used for the benefit of teachers (providing additional incentives or allowances). Parents feel that as their children work on the farm, children should be the ones benefitting from the proceeds.

Within Uyui Better, teachers discussed children from the community as inherently lacking in the capacities and capabilities required to achieve educational success. All educational failures were attributed to the family by teachers with little to no responsibility for educational failures taken by teachers themselves.

> "… *even if you try to help them in the class, no improvement occurs. So we assign them to do other activities like farming so as after completing their primary education they can even try to use farming knowledge for their life.* "

> "*There was a certain parent who came to complain that his standard 4 child does not know how to write but after we made some follow ups we found that even that parent does not know how to read while he has completed standard 7, so it is the family problem.*"

Teachers discussed the joint responsibility of the school and community for ensuring attendance of pupils at school. Teachers should know whether a child attends school, whilst the community should report children they see outside school during school hours to the teachers. Parents should be responsible for ensuring that their children actually go to school. The community leader thinks that he too has a role to play in ensuring that the community has an environment conducive to education and that the school has decent facilities. He also thinks that parents have a responsibility for following up on whether their children attend school and learning was achieved. Parents felt that the educational achievements of their children were low and cited the constraints facing teachers as possible explanations. Teachers thought that the poor performance of girls was partly their fault, but also partly due to the many household duties girl children have to attend to after school hours.

Student absenteeism from school is described as worse during the rainy season when only 50 to 60% of pupils are in school. This is due to the fact that parents require the help of their children on farms during this time of the year. Long periods of absenteeism from school are reported to discourage pupils when they return. Outside this period attendance reportedly rises to 80 to 85%. Children are reported to drop out of school in standards 6 and 7. The reasons cited for girls are relationships or family pressure to get married, whilst boys tend to drop out because of peer groups and alcohol consumption. Children from nomadic families are said to struggle because boys have to take care of livestock and girls are encouraged to get married at an early age whilst children from poor families struggle because they don't have uniforms, shoes, and exercise books which has a discouraging effect on them. In the earlier grades, problems with attendance, outside of the rainy season, is caused by the long distances children have to walk and the facilities at school. In the past, the school did not provide food to children even though children were unable to walk home for lunch as many live five to six km from school. The introduction of school feeding at this school is universally credited with having improved attendance by all stakeholders.

The community leader conceptualises his role as being responsible for ensuring that the relationship between the school and community is good and thinks that this is the case here which makes Uyui Better different from other schools in Tanzania. Nonetheless, he feels that the community is unaware of the importance of education. He attributes this to their own lack of education which he views as the reason why they fail to send their children to secondary school, even if they pass the exam. Despite the Millennium Village project and other awareness raising meetings he thinks that the importance of education is still not universally acknowledged and if it is, then only for primary school education. He attributes this to the high levels of unemployment which leads parents to think that it is unlikely that their children will find a job and secondary education is hence perceived as an unnecessary expense. He views the issue of teacher housing as negatively impacting on his ability to raise awareness of the importance of education in the community, as teachers don't live around the school compound and are demotivated and hence not able to help with raising awareness about the importance of education. In

his opinion, the head teacher should always live in the community which would enable her to form close ties with the parents.

The school committee does not feel that it has adequate skills to deal with its responsibilities of financial oversight and management and a lack of financial resources is perceived as the main challenge the school faces. IGAs are in place and crops are cultivated and sold to buy basic items for the school, such as chalk. However, yields are low and crops have been stolen from the school farm in the past. Nonetheless, IGAs are viewed as the best hope for improving the financial situation of the school. The community leader is of the opinion that the community is good at supporting the school, including contributing in-kind (material and man power to build a building) and allowing their children to work on the school farm after school hours.

M.2.1.4   Summary of selected key findings

At Uyui Better three main factors affect the amount of instructional time pupils receive: (1) teachers live far away from the school and travel long distances to come to school, often arriving half way through the morning, (2) late payment of salaries necessitates trips by teachers to the districts and reportedly also leads to the adoption of other income supplementing activities, and (3) pupil absenteeism during the rainy season reportedly reaches a staggering 40 to 50 per cent. If taken together, these three factors arguably have a great impact on the instructional time provided at Uyui Better.

At Uyui Better the former and current head teacher are engaged in a power struggle over who is in charge of running the school which is further exacerbated by gender dynamics and the divergence in qualifications between the two candidates. At the same time, different perceptions of the boundaries of corporal punishment, the delays in building teacher houses and acceptable levels of teacher absenteeism and teaching quality provide ongoing sources of conflict between the school and the community.

M.2.2 Uyui Typically Performing Primary School (Uyui Typical)

Uyui Typical is the largest school of the nine case study schools (940 pupils). It has the second largest class sizes (72) and the highest pupil-teacher ratio (85). The standard one to three pupil absence rate is high (55%) and is higher for boys than for girls. The head teacher is female and has been in post for six months. The school performs typically, with some improvement in recent years. However, it is not close to achieving gender parity.

Community members in the associated ward cultivate tobacco and have benefitted from economic development activities. There is no health centre in the ward and community members travel to surrounding villages to access most social services. The marketing of tobacco and access to markets more generally is a problem which negatively affects the living standard of the community.

M.2.2.1   Teacher capacity, morale and behaviour

At Uyui Typical, teachers' performance and morale is limited due to a lack of available teacher housing, late and low salary payments and high pupil-teacher ratios.

The school currently only provides housing for the head teacher and the houses teachers can rent are in poor condition. However, unlike the other teachers in the district teachers are at least able to rent houses close to the school. Nonetheless, teachers lack motivation and do not like the environment they live in due to a shortage of water and a lack of hospitals and transportation. The late and low salary payment is reported to drive teachers into debt and teachers are demoralised by multiple trips to the District Office to repeatedly fill in the same forms in order to claim their salary. This also

impinges on instructional time at the school which is further exacerbated by the fact that all teachers engage in income supplementing activities outside of their school commitments.

Teachers' frustrations derive from feeling poorly respected by community members because of the low educational achievements at the school. Whilst parents complain about the quality of teachers at the school and their inability to explain concepts to pupils, teachers report that they spend time helping the children who are struggling and attempt to build relationships with their pupils so that they can understand the problems each student has.

> *"For those who cannot read, you make a friendship with the child and finding extra time to learn. For example if the child cannot count I tell him/her to come with sticks that will help to count easily like 1+1 =2, so we count together. When you call, the child is not afraid of the teacher."*

The high pupil-teacher ratio was observed as leading to low morale and performance of teachers, with teachers struggling to manage pupils in large classes. Teachers mention the need for training, but have been to training recently and tended to focus on needs within the schooling environment as constraints to fulfilling their roles.

More highly qualified teachers perceived being deployed to the early years as an insult and equated skill with subject knowledge, rather than pedagogy. The community also described the teachers deployed to pre-primary, standard I and standard II as *'the incompetent ones'.* The teachers acknowledged the language complexities for teaching in the early years, but thought that this could be solved by ensuring that children use Kiswahili in school and their native language at home.

M.2.2.2 School leadership and management

The SC stated that there is a good record of teacher attendance, however teachers discussed an informal arrangement whereby they cover for one another if a teacher is not at school, or not attending class. In the past, the head teacher has provided financial incentives to teachers from IGAs in an attempt to motivate them to teach remedial classes in order to improve the performance in standard four and seven. Pupils were asked to stay at school during the holidays to get assistance and cover material not yet fully understood or covered. However, some parents did not understand why this was necessary and others refused to contribute financially to incentivise teachers.

The SC has decided that the remaining crops from the IGA would not be sold, but used to provide porridge for the children. Communication around these activities seems to lack transparency, as parents are unaware of this decision. Parents report that their children do not receive enough food at school and express frustration at not knowing what happens to the food the school cultivates. Parents feel that as their children are providing the labour making the income generating activity possible, they should benefit directly from the profit made. This is a common viewpoint expressed by parents across most case study sites.

As with Uyui Better, the head teacher at Uyui Typical used reporting up to the WEC as her only source of authority within the school. Unsurprisingly, teachers discussed the WEC and not the head teacher as their first line of accountability. The head teacher stated that she finds it hard to manage the teachers as they fail to respond to instructions and at times, teachers only teach five minutes in a period. The head teacher attributed this lack of authority to the fact that her salary is the same as that of the other teachers. Despite the WEC being perceived as the disciplinarian by both the head teacher and the teacher, he was reported as only having visited the school three times since the head teacher started in the role six months ago.

The relationship between teachers and district officials is difficult and fuelled by suspicion and frustration. As in Uyui Better, teachers express frustration at having to repeatedly fill in forms for salary claims and not being paid. In addition, teachers report that if they do receive training, officials keep to themselves which teachers perceive as due to unwillingness by officials to respect them. Teachers believe that officials think that they are better than the teachers they manage.

> *"We are not respected at all. I remember one day I went to the district education office to register the school equipment I had purchased and one person came to me saying 'you bought pens! I know you're going to give them to your own children. Do not come here pretending you have come for the good cause while we know your missions are to benefit yourself..' I felt very insulted."*

M.2.2.3   Relationship with the Community

Parents do not feel that their children's knowledge of Kiswahili is a problem and rather attribute the children's lack of understanding to hunger, a shortage of desks and children's tiredness after the long distances they walk to school. While teachers at Uyui Typical think that the community fails to value education, the community makes numerous in-kind donations and agreed to send their children to school on Saturdays for remedial classes.

The community state that they want the teachers to have access to housing in the community but felt that it is the responsibility of government to provide them.

> *"That is not our responsibility to build houses for the teachers while teachers have been brought here by the government…"*

The head teacher explains that the relationship between teachers and parents was not strong and when called to attend meetings at school parents refused to come. Tensions also exist between the community and teachers.

As in the other schools in the district, the administering of corporal punishment is a source of conflict. Parents report of instances where teachers have gone as far as kicking and punching children. Recently there was a case where a parent had a fight with a teacher and the case had to be solved by the village council, because the SC chairman has failed to solve issues concerning excessive corporal punishment despite repeated complaints by parents. The head teacher confirmed that children sustaining injuries while being punished by teachers was a common occurrence.

> *"If it is a small injury we make excuse to the parents and tell him/her it was a mistake and we take the child to the hospital."*

The community feels that "the teacher is a mirror of the society" and children mimic teachers behaviour and mannerisms. As a result parents think that the use of abusive language and inappropriate dress code among the teachers has a negative effect on their children.

Parents say that they would like to be involved in the academic development of their children and describe updates on their children's progress as the happiest moments at school. Parents also claim to cross check the exercise books of children to see what is being taught at school which shows that they conceptualise this as part of their responsibilities.

M.2.2.4   Summary of selected key findings

Teacher morale at Uyui Typical is low due to a lack of teacher houses, low and late payment of salaries and high pupil-teacher ratios. In addition, teachers feel that they are not respected by the community and government officials which reportedly further affects morale and motivation. As parents

acknowledge that teachers function as role models for their children, they are unhappy with the dress code of some teachers and the harsh and abusive language used. At the same time, the head teacher struggles to manage teachers and attributes her lack of authority to the fact that head teacher salaries are not automatically higher than teacher salaries.

The relationship between the school and the community is difficult and the village council has been called in to solve conflicts because the SC is perceived as unable to mitigate between teachers and parents, especially around questions of corporal punishment and subsequent injuries of children. Communication and transparency around what happens with the proceeds from the school's IGAs is not effective and parents are either unaware of what the money is used for or struggle to understand why incentivising teachers to teach remedial classes is necessary and given priority over providing food to the children at school. Parents identify hunger as affecting their children's ability to learn.

M.2.3 Uyui Lower Performing Primary School (Uyui Lower)

Uyui Lower is a large school (687 pupils) with large classes (69) and a high pupil-teacher ratio (62). The standard one to three pupil absence rate is relatively low (28%) and is higher for boys than for girls. The head teacher is male and has been in post for one year. Uyui Lower is not performing well compared to other schools in the district. In previous years, Uyui Lower was given a black flag for performance. Nonetheless, improvement in recent years has been weak, but the school is closer to gender parity than other schools in the district.

The majority of the community members cultivate tobacco and the community leader reported that this yielded a relatively reliable source of income. Key stakeholders indicated that the infrastructure in the area is relatively good and a World Bank water scheme has improved agricultural development, whilst the Millennium Village Project was perceived as having had a positive impact on the school and community. The school has access to tapped water, but due to limited availability this is only used for the purpose of providing drinking water.

M.2.3.1   Teachers capacity, morale and behaviour

At Uyui Lower the majority of teachers showed a strong commitment to professional collegiality, even though some lacked morale. While teachers were frustrated with the schooling environment, mechanisms to support one another to fulfil their roles more effectively were observed. Teachers at Uyui Lower described professional development strategies including 'having a visitor in the classroom', meaning that another teacher is invited into the class to observe the teaching and learning strategies; and to provide feedback to the teacher. In addition, teachers informally discuss the challenges they face in conducting their teaching. Many of the teachers describe their love of teaching based on the intrinsic value of education.

As with other schools, teachers described experiencing a gap between their expectations and experiences of teaching. However most teachers at Uyui Lower focussed these discussions on classroom practice, child centred learning and the challenges of using participatory methods. Some teachers were less well motivated by the intrinsic desire to teach and rather were motivated by the fear of pupils telling their parents that teaching is not taking place at school. Other teachers being drunk at school was identified as an area of concern by the majority of teachers.

While most of the teachers are clearly motivated, they still engage in income generating activities outside of their roles, indicating that teacher absenteeism is unlikely to be attributable to low levels of teacher morale alone. One teacher described this in terms of the impact on his performance, stating *'the teachers' environment will always destroy their proficiency'.* The SC described this situation as an impasse stating that until teacher salaries are raised, other incentives will not be able to fully resolve the issue of teacher performance. Conversely, the head teacher stated that training would be better

than salary increases as '*teachers' minds are not settled and this cannot be treated with salary'.* At this school teachers support one another financially when salaries do not arrive on time. However, the SC stated the late teacher pay is not a problem in Tabora, a statement contradicting the findings from the other two schools in the region. Teachers regularly lose instructional time to organisational issues at the District Office. Checking on the status of salaries often takes time and the deployment of teachers to schools was described as an unclear process whereby teachers were expected to report to several offices without previous warning – a process which often takes several days. Teachers described the loss of income due to having to follow up on administrative issues and felt that this was more important than the level of the salary itself.

A number of issues affect teachers' ability to do their job. Teacher houses are in disrepair, which leads to absenteeism and late arrivals during the rainy season. Some teachers discussed sleeping in the classrooms at school as they are more comfortable there than in their homes. Language and curriculum were two areas that Uyui Lower teachers found particularly challenging. Pedagogies to support the teaching of Kiswahili in the early years was an area of concern and the introduction of new curriculum areas that were not viewed as relevant for rural living. For example, one teacher questioned the value of teaching a child how to turn a computer off and on, without access to a computer in the classroom or anywhere else in the community. Furthermore, the teaching load was perceived to be too high and teachers newly out of college felt eager to employ new teaching strategies but felt stifled by more experienced teachers who were not open to change. Teachers also felt that being asked to teach subjects that they have not specialised in, leads to poor teacher performance.

Most of the teachers at Uyui Lower are less constrained by motivational factors than teachers at other schools. The teachers named a broad range of topics they would like training in and it is likely that the provision of training would further motivate Uyui Lower teachers. However, teachers are critical of the training they have received in the past stating that it was on topics they were already familiar with instead of on curriculum changes. Another are of concern for teachers is if training is provided to one staff member who is then expected to train the other teachers.

> *"Me and my colleague went to be training on how to teach standard one, but when we arrive there, they only trained my colleague with the intention that when she gets back, she can train all of us, except there was nothing new on what they trained us. It was everything we already knew. Imagine, one person is been trained to come and train the others. Does this make any sense?"*

M.2.3.2   School Leadership and Management

Uyui Lower had a serious leadership crisis prior to the appointment of the current head teacher. Teachers described undertaking their duties under the previous head as traumatic and *'like being on fire'*. In addition, financial resources were embezzled leading to conflict with parents and the wider community. Eventually the teachers wrote to the District Office for support. The community also got involved and requested the removal of the previous head teacher. The WEC at Uyui Lower described the school's teachers as motivated after the school had previously been given a black flag for performance in examinations. However, the teachers' relationships with the District Office remains difficult. Teachers described being insulted at the District Office and being treated like children. It was unclear if this was the result of the conflict from the previous years. Either way, the relationship between teachers and officials seems to have sustained permanent damage.

> *"How the education officers reacted we were discouraged. We felt we had no place to take our complaints……we were embarrassed, we were insulted, and he send a group of people. Since then when you have an issue to do at the education office you go and gaze at the window to see who is inside and you know whom you can call out to come and help you outside."*

The current head teacher conceptualises his role in terms of supporting and managing teachers, being a conduit between teachers and parents, guiding the community and providing academic leadership. The head teacher advocates for the teachers to come together *'for one common purpose'* and has sought to institute in-service training within the ward, stating that the schools should learn from one another rather than wait for the district officer to come and provide education. At Uyui Lower the head teacher eases the tensions between teachers and parents by asking teachers to bring requests for corporal punishment to him prior to hitting children, asking community members to direct questions to him before complaining to teachers and asking pupils to meet with him prior to talking to their parents about a teacher. Building positive relationships with pupils through this process was the deliberately stated aim. Furthermore, the head teacher discussed one of the key components of his role as *'maintaining the standard of academic performance',* checking teacher's preparation schemes, lesson plans and observing teaching. In addition, the head teacher provides incentives and petrol to the WEC to deliver training to his teachers.

M.2.3.3   Relationship with the Community

The main cause of tension between the school and the community is the behaviour of teachers. Parents stated that the teachers spend time on their phones during class and fail to dress appropriately. Parents compared the teachers at the school to the teachers that they experienced when they were at school. Parents showed empathy for the environment that teachers work in and understood that the state of teacher housing was insufficient. Teachers also showed an understanding of the workload of pupils outside of school and adjusted their teaching strategies to accommodate this. For example, some teachers indicated that they allowed some pupils to do their homework at school, as long as they arrived before classes started. However parents also discussed the presence of 'mean money' at the school, whereby parents can pay teachers to ensure their child receives attention within the classroom.

Another source of conflict at the school is the administration of corporal punishment. The rules that stipulate how and when corporal punishment may be administered by teachers has been changed in the past. This has led to a situation where the community perceives corporal punishment as abolished, whilst teachers insist that changes merely provide for stricter rules for when corporal punishment can be administered to children. The school therefore continues to administer corporal punishment despite protests by parents. In the past this issue caused tensions between parents and teachers leading to claims that parents hated teachers. Parents are divided on whether corporal punishment is acceptable and teachers report avoiding punishing the children of parents who are likely to come to school and complain.

The SC reports that the one difficulty in the relationship with parents is that they sometimes do not come to meetings and are unhappy when called to school to discuss their child. The relationship between the SC and teachers is described as good. However, the head teacher complained that parents raise grievances in annual meetings rather than approach a teacher directly and give him a chance to explain him or herself. The community leader agrees that unhappiness with teaching should be addressed by approaching the teacher and head teacher directly. However, the community leader explained that only educated parents can check on pupils' progress by checking the exercise books. Other parents have to come to school and rely on the school to provide an update in order to learn about their child's academic progress. These parents are thus restricted in their ability to provide support on monitoring the teaching at school.

Parents report that in previous years they were unhappy with the situation at the school as they perceived their children as coming to school to fetch water, collect grass to build teachers houses and other non-academic activities, all during class time. The change in leadership at the school has had a positive impact on the community. Prior to the change the community felt that teachers were misusing funds and food at the school. A new regime of absolute financial transparency was introduced by the new head teacher which ensures that the community is aware of everything happening at the school.

This has motivated the community to be more involved in helping the school and ensuring that the school has enough food to feed the pupils. Even members of the community who do not have children at the school contribute three buckets of maize to the school, something usually only done by parents, not all members of the community. The community leader attributes this to a sense of ownership of the school by the community.

> *"This school is our school, whether you have a student or not it is your school so any problems at school you have to solve."*

Additional income for the school is derived from income generating activities. The school cultivates maize, sunflower, millet, groundnuts and sweet potatoes and sometimes pupils work on the farms after school hours. However, the school committee reports that this isn't the case on a daily basis. Proceeds from the IGA is used to buy teaching aids, chalks and other necessities. Some food is retained for feeding the children. In addition, the school is starting a brick making business for which it employs labourers from the village.

M.2.3.4   Summary of selected key findings

The teachers at Uyui Lower are comparatively motivated, even though the environment in which they live and work frustrates them. Teachers have developed mechanisms for coping with the situation and provide support to one another at school and at home. Nonetheless, the poor state of housing, the need to follow up on administrative problems at the district level and to supplement income through other jobs or activities is widely acknowledged as decreasing instructional time and affecting teacher performance. However, the SC perceives this as an inevitable outcome of the situation and is understanding of the need of teachers to deal with these constraints. Parents largely agree, but criticise the behaviour of some teachers at school, especially the conducting of phone calls during class time. Interestingly, parents are clearly viewed as able to hold teachers to account and seem well aware of what is happening at school.

Previously, Uyui Lower faced a leadership crisis which teachers were unable to resolve and was only addressed by the district after parents involved the village council. Whilst the new head teacher is well liked and has a very comprehensive view of his roles and responsibilities that includes managing and supporting teachers, leading, and mitigating between the school and parents, the relationship between teachers and the district continuous to be strained as a result of the past. A commitment to total transparency around management and financial transactions at the school has restored the confidence of the community in the school and reportedly increased the willingness of parents to support the school.

M.3   Kishapu, district Shinyanga
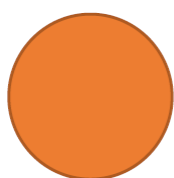
Kishapu, Shinyanga was selected as one of three qualitative case study districts on the basis of the available quantitative data on literacy, school resources, social and economic resources and the GPI at the lower secondary level. Kishapu is a district performing typically, despite higher school level resources and typical social and economic resources. The percentage of girls enrolled at lower secondary level in Kishapu is significantly lower than in the two other selected districts. The vast majority of Kishapu residents are subsistence farmers and cattle herders. The district is open to economic shocks on the basis of unpredictable and insufficient rainfall, resulting in famine and malnutrition.

Kishapu is the only district where some of the head teachers had attended in-service training during 2012 to 2013. The head teachers of Kishapu Better and Kishapu Lower Performing Primary School

had both attended government-funded training with the former also attending some donor-funded training.

Three schools were purposively sampled in the Kishapu district.

M.3.1 Kishapu Better Performing Primary School (Kishapu Better)

Kishapu Better is a fairly small school (319 pupils) and has below-average class sizes and pupil-teacher ratios (46 and 53). The standard one to three pupil absence rates (45%) are above average compared to other case study schools and higher for boys than for girls. The head teacher is male and has been in post for two years. Kishapu Better is located in the Lalago district and was selected as a school performing reasonably well, compared to other schools in Kishapu.

The community members are farmers and cattle keepers and the majority of pupils are kisukuma speakers. The low cotton price has negatively affected this community which relies heavily on cotton cultivation. The school is located several kilometres from the town and relatively hard to reach. Several grades have to share classrooms leading to reduced class time. Water has to be fetched from a source four kilometres away from the school.

M.3.1.1 Teacher capacity, morale and behaviour

Kishapu Better teachers described an intrinsic joy in explaining and having pupils understand concepts within the classroom and seeing pupils go on to secondary school.

The challenges that teachers at Kishapu Better face that affect morale include low salaries and a lack of teacher houses. The SC has put in place measures to motivate teachers which include giving gifts and awards to teachers who perform well, whilst the head teacher provides monetary incentives and believes that money is the most effective way of motivating teachers.

Constant syllabus changes coupled with a lack of training on how to adapt to changes such as the inclusion of subjects teachers had not specialised in were cited as a key challenge to educators. The lack of clarity on curriculum content, assessment processes and the correct textbooks the teachers should use were further mentioned. The older teachers are mostly affected by subject changes that include teaching about new technologies. Some of the teachers expressed that if they were examined on new computer subjects they would fail.

> *"…if you give me the exam of TEHAMA I will fail because I know nothing of it. There is the use of computers in that subject which I do not know and also there is a new thing called laptop. I have never touched it but there is a picture of it on the TEHAMA book and I am asked to teach it. How can I manage?"*

Teachers report that when inspectors come to the school they ask questions about why older teachers are not teaching the new subjects like TEHAMA and they have been told that teachers require training in order to be able to teach these subjects. Nonetheless, training has not yet been provided. All teachers agreed that they required training when new material is added to the syllabus and curriculum. The number of new textbooks that are being introduced is also perceived as a problem by many teachers. Teachers are worried by the fact that schools in the same district are using different textbooks to teach the same subject.

> *"When he [a teacher] went to pay a visit to a neighbouring school he found different books. He was very confused and he did not know which book to use. The same case applies here. I teach mathematics and I have four Mathematics books. One is published by action, others by Ben, Mture and Oxford. I am very confused about which book I should use."*

As with all of the nine case study schools, teachers engage in income supplementing activates to increase their income. All stakeholder groups at Kishapu Better acknowledge that teacher absenteeism from school is not formally acceptable, but feel that the low level of salaries and frequent late payment de facto force teachers to look for ways of supplementing their income. In relation to the impact on teacher attendance, the community leader stated *'what can they do? They can't do two things at the same time'.* Parents reported that teachers often do not come to school. As a result parents complimented the teachers on their entrepreneurial skills rather than their teaching abilities and felt that if teachers were provided with adequate shelter and water and were still not teaching, the community would be in a position to criticise them. Under the current circumstances, teacher absenteeism and the pursuit of other jobs or income supplementing activities was accepted within the community as an undesirable but unavoidable. As with pupil attendance, teacher attendance was reported to drop dramatically during the rainy season when teachers concentrate on their farms and not on teaching at school.

A major cause of loss of instructional time at Kishapu Better was a shortage of classrooms and teachers. For example, one teacher is required to teach pre-primary and standard one and two.

> *"We don't have enough teachers because a teacher who teaches standard one and two is the same teacher who teaches the pre-primary pupils. So it is a bit hard for him or her to maintain the time to teach all the classes, which means he can attend 30 periods in a week."*

At Kishapu Better, the question of how Kiswahili should be taught in the early years was raised by teachers and the community leader. The teachers deployed to the early years do not speak Kisukuma and therefore, a full immersion program is undertaken where Kiswahili is the only language spoken in the classroom. The community leader at Kishapu Better understands that special skills are required for teaching Kiswahili in the early years of schooling. As a result, pre-primary teachers are respected within the community – a marked difference to the situation at other case study schools. Parents generally felt that their children's limited Kiswahili skills did not negatively affect their ability to learn at school.

M.3.1.2 School Leadership

The head teacher at Kishapu Better described himself as strict and perceived his role as mainly involving following up on teacher and pupil absenteeism, administrative duties and liaising with the community. The head teacher does not think that the school has a problem with student attendance, even though the rate of pupil absenteeism was higher than at other case study schools. The community leader and parents discussed that the previous head teacher was disliked by teachers because he closely watched their attendance record.

> *"Some time back, pupils used to just roam around all day. Teachers only focused on their businesses. The head master held a meeting with teachers and asked them to start doing their jobs as required of them. That was the beginning of teachers hating the principal. They spread bad rumours about him and they even tried to beat him up. They also requested for his transfer. That was in March if not April. This (new) headmaster is very understanding. That's why I am saying these teachers should be transferred because they don't care about our children."*

Interestingly, the new head teacher who has been at the school for two years, perceived the previous head teacher as problematic and causing division between the teachers. The head teacher identified this is as the reason why he was transferred to Kishapu Better and describes himself as having a good track record at turning schools around. Parents confirms this assessment of the situation, claiming that they have seen a positive change since the arrival of the new head teacher.

"At least when this new school principal came, he brought some emphasis with him. All pupils did in school was play, fight and make noise but that has reduced since this new teacher came."

Among other changes initiated by the new head teacher is the directing of school funds towards essential development projects. For example, the school previously did not have toilets for its pupils. However, the head teacher discusses not having sufficient funds for running the school, even though the school occasionally receives additional donations from government. Recently shoes were delivered for the children, but not all children received shoes and most were broken after two months.

The WEC has been to visit the school three or four times this year. The head teacher describes feeling supported by the WEC and sees him as a partner who shares the burden and responsibility of making sure that the school improves. Successes and failures are perceived as jointly owned by the WEC and the head teacher. However, like all other WECs in the district he struggles with a lack of funds to cover his travel costs. The community leader reported that the WEC has never attended a village meeting, but acknowledges that the head teacher does attend and provides information about the school and its income and expenditure. Despite this disclosure of the financial situation at the school, the community leader still feels that he does not know what happens to the financial resources the school receives.

M.3.1.3 Relationship with the community

The community were extremely unhappy with the behaviour of teachers within the school at Kishapu Better. The relationship between teachers and parents constituted the greatest area of conflict at the school.

The parents feel that the head teacher is the only person who does a good job and want all teachers to be removed from the school.

"We want all these teachers except the head teacher to be removed and have new teachers. They have been here for a long time."

One parent said there was a time when all the teachers refused to teach and the head teacher had to teach all the subjects in standard seven. The children passed the exam and are now in secondary school. This achievement seems to have greatly contributed to the generally high standing of the head teacher within the community. Parents also report visiting the school to discover that all teachers are absent from the school and children are left unsupervised. In addition, teachers were viewed as forcing girls to undertake domestic duties in their homes.

Allegations of sexual relationships between teachers and pupils were frequently made by parents. One teacher in particular was accused of sleeping with female pupils. When asked to describe his worst moment as a teacher, a male teacher described a scenario at a previous school he taught at where he was accused of having a relationship with a student. He describes the process as unprofessional and was angry that he was only allowed to provide his side of the story towards the end of the process. He feels that the lack of proper procedures for dealing with these types of accusations can lead to the destruction of the reputation of a teacher.

Parents acknowledged that the living conditions of teachers could be improved by building better teacher houses and providing clean water. Parents feel that this would enable them to attract better teachers to their school and hold teachers to account for their teaching performance.

"If teachers were provided with good shelters and water and still would not teach, we would strangle them but for now we have no reason to."

The village council and community in general is not supportive of the school. The SC struggles to engage with parents and the village council. When information is provided to the village council it is not addressed in a timely manner and information usually reaches the community with great time delays. The head teacher indicated that the community trusts the SC with financial records and transactions, whilst parents claim that their financial contributions for the construction of teacher houses have been misappropriated. The SC does not receive adequate support from parents for new projects planned at the school. This seems to be due to past problems at the school and the lasting negative impact this has had on the relationship between the school, the SC and the community. Teachers are more understanding than the SC of parents' reluctance to support the school financially and attribute this to the general level of poverty within the community.

> *"The government tells parents to contribute money for making desks, but the parents who are told to contribute money for the desks if you decide to pay a visit at their home you will find she/he does not have even stool. Therefore even the importance of a desk is not known. Yet again they are told, the school does not have toilets you should go and dig while he/she does not have a toilet at her/his home."*

M.3.1.4 Summary of selected key findings

At Kishapu Better, teachers described an intrinsic joy in explaining and having pupils understand concepts within the classroom and seeing pupils go on to secondary school. Nonetheless, a lack of teacher houses and low salaries negatively affect morale. In addition, teachers struggle with the changing syllabus, availability of different textbooks and a lack of training provided. The SC tries to incentivize teachers with awards and gifts, but the head teacher provides financial incentives as he thinks that money is the only motivator for teachers. Teachers engage in income supplementing activates to increase their income – a behaviour largely seen as inevitable by stakeholders given the working conditions of teachers. Nonetheless, the loss of instructional time and teacher absenteeism is a source of great conflict with the community.

The relationship between teachers, parents and the wider community is very strained. The community wants all teachers removed from the school and is primarily interested in improving the working environment because they think it will help attract better teachers. The current teachers are described as lazy. However, the head teacher is viewed as very hard working, a good teacher, and frequently attends village meetings to report about events at the school. Nonetheless, the relationship with the village council is poor and requests by the school to relay information to the community are either ignored or passed on with great delay.

M.3.2 Kishapu Typically Performing School (Kishapu Typical)

Kishapu Typical is a relatively small school (373 pupils) with a relatively low pupil-teacher ratio (37). Class sizes are just below average (53) compared to the other case study schools. Standard one to three pupil absence rates are low (17%), but there is a sizable difference between boys (26%) and girls (10%). The acting head teacher is female. She has been at the school as a teacher for seven years and has acted as head teacher for the past 10 months. The school was selected as a typically performing school, with little improvement over recent years.

Community members are predominantly nomadic cattle herders or farmers and cultivate cotton, millet, sweet potatoes, rice and groundnuts. Absenteeism is high due to cattle herding and cultivating activities, especially during the rainy season. The community was described as 'extremely poor' and children often travel long distances to reach the school. Hunger during school time is wide spread due to the general levels of poverty in the community and the lack of a school feeding programme.

M.3.2.1   Teacher capacity, morale and behaviour

As in other case study schools, teachers experienced a 'disconnect' between their expectations of the profession and their actual experiences. Many of the teachers expressed disappointment in the government for placing them in areas with poor infrastructure and no houses for teachers. The teachers feel that the government and education officers do not care about them.

> *"When I arrived here in 2010 there were neither teacher houses nor toilets and bathrooms for teachers. Me and my family had to use pupils' toilets. For a kitchen I had built a fence out of grass, as you can see(points at the school store which he has turned into a house for himself and his family)."*

The teachers also claim that government has not made any new developments and improvements to facilities at the school since colonial rule. These perceptions lead to frustration and extremely low levels of morale in the teaching staff.

The school suffers from a shortage of desks and classrooms. Pupils habitually take lessons outside and study under trees and the teachers are concerned about the fact that the children have to learn outside, even during the winter months. The head teacher echoes these concerns and explains that pre-primary and standard two classes take place outside, whilst some standards or multiple streams have to share a classroom. This is likely to negatively impact on instructional time and impinge upon teachers' ability to interact with pupils. A further factor that leads to loss of instructional time is the delays with which parents return their children to school after the holiday period. Some children return to school two weeks or a month after school has resumed. Teachers struggle to find a way of addressing this problem and are reluctant to move ahead with teaching if the majority of pupils aren't at school. In the past, the head teacher dealt with student absenteeism by reporting pupils to the district which would then take parents to court. However, the village council intervened and defended parents and cases were allegedly dropped without any explanation. Teachers perceived this to be the case due to political interferences and reported being demoralised by this outcome.

The teachers at Kishapu Typical have a good working relationship and support one another in cases of knowledge gaps on instructional material or fill in for one another in cases of absence due to illness. However, there is shortage of teachers and as a result the teachers have to teach a lot of subjects across various grades.

Like most other teachers in the district, teachers at Kishapu Typical perceive the available textbooks as being inadequate for supporting teaching. Textbooks are described as incorrect, contradicting one another, going into different levels of depth on a subject for the same standard and not in line with the curriculum. In addition, teachers complain about additional teaching responsibility for subjects they weren't taught at teacher training colleges. This is a particular problem for older teachers who have not received training to keep up with the changing syllabus.

> *"There are big changes caused by science and technology such as there are new subjects for example TEHAMA (ICT), also the use of computer and modems. If you tell me to use these things, how will I manage while I have graduated college a long time ago? So it is important for us to get trained on these changes."*

The head teacher suggests in service training as a way of dealing with this problem. However, the different teaching methods used by teachers, knowledge of the syllabus, and training received reportedly causes problems beyond the school at the district level where the WEC may disagree with teachers, as he too would have gone through teacher training college at a specific point in time. This suggests that WECs also require training on curriculum and teaching practices changes. Teachers

suggested that WECs, the DEO and inspectors liaise more closely with teacher training colleges to deal with this source of conflict and confusion at schools.

The Kiswahili language skills of standard one and two pupils is a big source of concern at this school, and one that unlike many other schools they deliberately try to address by adapting the staffing policy. Teachers who are qualified to teach pre-primary, but cannot speak the local language are not assigned to teach this class which is instead taught by teachers with longer service records at the school. However, whilst this addresses the language barrier between teachers and pupils it also means that latest lessons learned in pedagogy and on how to teach pre-primary are not applied. Teachers clearly feel that being able to communicate with pupils is more important.

> *"We have staff meetings and decide who teaches what subject by considering academic qualifications, willingness, and experience. For example you cannot give the teacher who doesn't know Kisukuma to teach pre-primary. So those who are long service teachers are the ones who teach pre-primary pupils or standard one. Because children don't know Kiswahili and they have to learn here at school."*

As in all case study schools, low promotion rates and salaries contribute to the lack of motivation experienced by teachers. Teachers reported that their salaries are too low to live on and that they hence have to ensure that they spend some of their time on other income supplementing activities.

> *"I think if a teacher can be given enough salary it will relieve them from stress. Enough salary will help a teacher to reorganize. Otherwise you can find a teacher doing other small business like carrying passengers on his motorcycle to increase income. With enough salary a teacher can agree to work from 6am to 6pm as he/she knows that at the end of the month he/she will be paid well. But for now teachers are focusing on increasing their income through raising chickens and having a vegetable garden."*

As a result, teachers are unwilling to work extra hours because they would rather use that time to generate additional income. Low levels of salary are thus contributing both to teacher absenteeism, low morale and diversion of time from teaching activities to income supplementing activities. The head teacher does not feel that teacher absenteeism is a problem and that absenteeism is mostly justified and the result of an illness or the collection of salaries. A permission book exists and teachers ask for permission to be absent. However, the SC says that teachers go to school just to sign the register and that this does not provide an accurate picture of teacher presence at school. The SC also feels that classroom absenteeism is a problem at this school and that the head teacher fails to hold teachers to account. The SC reports that excuses are made to cancel classes, as was the case when a teacher went to pick flowers as props rather than paint a picture of a flower. The SC is of the opinion that this constitutes unjustifiable loss of instructional time. As a result, the relationship between the SC and the head teacher is strained.

M.3.2.2   School Leadership

The new head teacher is described as a good leader who listens to advice given to her. The WEC has been to visit the school four times this year, but the school has to support him financially by paying for fuel in order to enable him to come.

In the past, the school experienced conflict around the administration of corporal punishment, but this has since been resolved. The previous head teacher was described as very harsh and parents were extremely unhappy. The new head teacher has decided that only she can punish pupils and requests signed permission from parents prior to doing so. However, some teachers feel that this makes them look weak and continue to hit children to assert their authority, claiming that the lack of respect displayed by pupils necessitates corporal punishment.

> *"The head teacher is the only one who is allowed to punish a student, but because of how todays' children behave, we just punish them. They are very disrespectful so if you wait for the head teacher to punish them, to them you will look as a statue."*

The head teacher mostly relies on alternative ways of punishing children such as tasking them with collecting stones and working on the farm. Some of the teachers also report looking for methods of punishment that suit the crime.

> *"You cannot beat or tell a child to collect stones as a punishment if he got the exercise wrong during class. Instead you should give him more exercises till he gets it right. You can even not let him/her have lunch break till he/she gets the right answers, I think that is the right punishment."*

Parents are not happy with the way the SC deals with issues relating to the use of school funds. They feel that the SC is not skilled to deal with financial matters at the school. Furthermore, teachers think that the procurement processes are tiresome and time consuming, especially if related to the purchase of stationery, like flip charts, chalk and other teaching aids. Teachers report that they write up a list of everything they need and the SC then approves the budget. However there are things the SC doesn't understand or knows the value of, allowing for collusion and fraud. The list is then given to the head teacher, who gives it to the WEC who sends it on to the district. After approval is obtained the school is sent a bank cheque. This process can reportedly take up to a month.

> *"A SC can ask what is stamp pin and a teacher tells them it is something which cost 200 000tshs and a flip chart costs 400 000tshs. They approve the budget thinking that is how much these items are and that they are important for teachers to have. And when the list reaches the district level, there is a person there who approves it believing, if the money is given he will also have a share."*

Teachers would prefer it if the head teacher were to be given the power to buy school materials without having to go through the DEO in order to speed up the process.

M.3.2.3   Relationship with community

The current head teacher has made a lot of progress in repairing the relationship between teachers and the community. The community leader describes teachers as having the characteristics of quality teachers and conceptualises the role of the head teacher as the educator of parents on the value of education.

One of the key challenges faced by the head teacher is the language barrier between parents and teachers. Only two teachers speak Kisukuma and all correspondence with parents is written in Kiswahili so the child then has to read out the letters to parents and translate it into Kisukuma. For teacher parent meetings, the head teacher has to rely on the SC to translate. Given the strained relationship between the SC and teachers this poses an additional barrier to effective communication. The community leader reports that an inability to communicate is often perceived as disinterest by the parents.

> *"There is the importance for communication. The parent may come to school without knowing how to speak Kiswahili. Maybe he is an elder one. So parents sometimes thought teachers don't care for them. So they [parents] have to learn and speak kisukuma to please parents."*

Another area of disagreement between the school and parents is around whether time at school should purely be spent on education. Parents are of this view and express unhappiness at their children fetching water and helping to work on teachers' farms during class time.

M.3.2.4   Summary of selected key findings

Kishapu Typical faces a shortage of classrooms and desks and classes are habitually conducted outside or with several standards or work streams sharing a room. Teachers use the low level of their salaries as a justification for their income supplementing activities and are unwilling to work extra hours because they would rather use that time to generate additional income. Low levels of salary are thus contributing both to teacher absenteeism, low morale and diversion of time from teaching activities to income supplementing activities. In addition to the confusion caused by the availability of numerous textbooks for the same standard and subject, teachers report that the different training received by teachers, head teachers and the WEC due to the different times at which they graduated leads to diverging believes around what is the correct method of teaching. Training to ensure that teachers and the WECs share the same understanding of pedagogy and have the necessary knowledge to teach all subjects would be of great benefit.

.
The relationship between the SC and the head teacher is strained, as the SC feels that teacher absenteeism and classroom absenteeism are a huge problem at the school and that the head teacher does not want to address this problem. The SC argues that the mechanisms in place to monitor teacher attendance are flawed and manipulated. Concerns about the ability of the SC to deal with financial matters at the school were raised by parents and teachers.

At Kishapu Typical, the existence of a language barrier is acknowledged by most stakeholders. The SC translates during meetings and teachers who have learned Kisukuma are used to teach pre-primary instead of specially qualified teachers. However, the community leader stated that the language barrier between teachers and parents is often misunderstood as disinterest by parents who fail to get an acceptable response from teachers when they come to school. As in many other schools, the use of corporal punishment is a divisive issue and the head teacher sought to deal with it by claiming a monopoly on its use. However, teachers still continue to use it, arguing that pupils would otherwise fail to respect them.

M.3.3 Kishapu Lower Performing Primary School (Kishapu Lower)

Kishapu Lower is a medium-sized school (477 pupils) and class sizes and the pupil-teacher ratio is average compared to the other case study schools (68 and 59 respectively). However, the pupil absence rates are the second highest of the nine schools (59%) and are higher for boys than for girls. The head teacher is male and has been in the post for six years. Like most of the community, he is from the Sikuma tribe. The school was selected as a lower performing school but with high improvement in recent years.

The majority of community members are either farmers - cultivating cotton, maize and groundnuts - or nomadic cattle herders from the Sikuma tribe. The community has been affected by recent low rainfall and often struggles to feed their families. The school has a water dam, but due to the recent droughts the water has mostly dried up and has to be fetched from a well further away.

M.3.3.1   Teacher capacity, morale and behaviour

Most of the teachers at Kishapu Lower have fond memories of their first years as a teacher. However, over the years they have become demoralised due to the multitude of challenges they face at schools they have been posted to. A lack of adequate housing has been a problem for all teachers since taking up the profession. Low salaries were also cited as affecting morale and motivation. Teachers generally report feeling overworked and underpaid.

> *"I do not concentrate on the government salary because I will never develop. I decided to do my own things."*

The teachers state that the poor infrastructure at school and the lack of housing provided make it difficult for them to adequately perform their job. A lack of proper toilets, offices, houses for teachers, classrooms, and a lack of teaching aids for information, communication and technology (TEHAMA) were described as the main challenges they face at the school. The head teacher added that teachers feel abandoned because the government does not send them for training. Even though teachers know and believe that Tanzania has a large body of female teachers, Kishapu Lower only has one female teacher. Though an extreme case, this is in line with the views held at other schools that schools in rural areas have more male than female teachers, as men are better equipped to deal with the harsh living conditions in rural areas – a view contradicted by communities who see female teachers as better able and more willing to adapt. All the teachers at the school feel that having only one female teacher, especially at a primary school, is a problem. Teachers believe that certain problems at school can only be dealt with by female teachers and these types of issues reportedly arise on a weekly basis. For example, male teachers feel that they cannot support a girl who is going through puberty and cannot check female pupils' during cleanliness inspections. Providing this type of support to female pupils is conceptualised by all stakeholders as the responsibility of female teachers who are meant to provide guidance and check all female pupils at school.

As is the case with the other schools in the district, the issue of having contradicting textbooks for one subject is a challenge for teachers. The textbooks are viewed as not aligned with the curriculum and the choice of five different textbooks for one subject confuses teachers who then often prefer to use the old books rather than new ones. As a result stacks of unpacked new textbooks can be found in the school office which teachers feel unable to use for teaching purposes.

> *"When teaching we have syllabus and content. You may have prolonged contents [in the textbooks] that are not in the syllabus. Hence teachers are confused."*

Furthermore, teachers feel that they are not included in decisions about syllabus amendments. One of the parents expressed concern about children doing well in the lower grades and on internal exams, but then going on to fail national exams.

> *"The most annoying thing is when a child is doing well in her/his studies and school exams but fails the final exam, this is very surprising. Maybe the curriculum used by the government for preparing final exams is quite different from the one used for teaching."*

The teachers at the school do not work well with the head teacher and the SC. This was very clear from the parents' analysis of the school. There is one teacher in particular who only comes to school twice a month and is perceived by parents as punishing children for no reason.

> *"He may come to school in August and tell the student to write the dates of the last month. How come he told the student to write last month`s dates while he comes today to teach."*

However, the SC has a contrasting viewpoint, as they think that there is no problem with teacher absenteeism and point towards a book where teachers write reasons for their absence. They feel that this book which was introduced by the head teacher demonstrates that teacher absenteeism is merely a perception problem.

Nonetheless, the SC concurs that teacher housing is a problem.

> *"To be honest our teachers have a very bad time because they don't have a place to live and we don't have houses for the teachers. The one, which they are living in right now they used to be classes for the pupils. So we decided to give them to the teachers. And for the time being there are only three teachers who are living in them even though we have eight teachers."*

The head teacher thinks that children are proficient in Kiswahili and that claims by parents and pupils to not be able to understand teachers are actually a way of asserting themselves over teachers.

> *"It's not that if you speak Kiswahili they don't understand you. They do but they are just forcing you to speak Kisukuma all the time so if you agree with them they destroy your aims that I am not speaking Kisukuma but Kiswahili. For example, now we have a village executive officer who doesn't know Kisukuma but he copes with them [the community]. They know Kiswahili but they just want to destroy your aims."*

Teachers agree that unlike in other schools, the children do understand Kiswahili. The problem is that they prefer to speak Kisukuma. As a result teachers report difficulties when trying to communicate with children. It isn't clear why teachers and the head teachers think this is due to unwillingness on the part of pupils rather than an inability to understand. However, when talking to elder community members the language barrier is perceived as being real by all actors.

M.3.3.2 School Leadership

The relationship between the school and district level is difficult on a number of levels. The teachers at Kishapu Lower report being unhappy with the education officers because they feel that they do not address their problems; and only visit schools every two to three years. The last time they saw them at their school was in February/March 2013. The SC also perceives support from the district level as lacking.

> *"For instance the building that we have, the unfinished ones, the community is the one who was involved in building them. They are unfinished because in Kishapu District before you ask for a project you need to show it by starting yourself first in order to be given help. For instance the District Executive Officer told us to start the building and he will later on finish it."*

While the community has held up its side of the bargain and started building the new classrooms, the DEO is has not provided support which has caused anger within the community and has made parents reluctant to contribute further. On the other hand, the municipality recently provided the school with ten extra desks which the SC sees as a positive sign. One SC member reports specifically joining the committee in order to apply pressure on the government to provide more support to schools.

Another development teachers are unhappy with is that government has changed the procurement procedures that schools have to follow to acquire books. Instead of giving the school money to buy books, they buy the books and send them to schools. The teachers feel that the books procured by government are not the right ones and when inspectors come to the school they tell the teachers they are not well prepared to use the new syllabus. Resources are being wasted on buying textbooks that the teachers do not use because – in their mind - they are not aligned with the syllabus and are perceived as confusing teachers on what they are expected to teach on.

Parents feel that teachers do not respect the head teacher and undermine his authority.

> *"The head teacher accepted that the children come [to school] at 7:00 am, but the other teachers tell the pupils to be at school at 5:30am or 6:00 am. They don't respect the head teacher's decisions".*

When the parents complained to the head teacher about pupils working in teacher houses, teachers failed to abolish this practise despite the head teacher's support of the parents' viewpoint. Parents feel that teachers do what they want at school, regardless of what behaviour is agreed as acceptable between parents and the head teacher.

> *"He is a nice man but there are some teachers who disrespect him."*

Similarly, when the SC reprimands teachers for punishing pupils, teachers react angrily. Parents strongly feel that there is a crisis of leadership at the school and that the head teacher and the SC fail to rein in the teachers and put an end to behaviour that negatively affects the relationship between the school and the community. The village council is called upon by the school to help deal with absenteeism of pupils, as it has the authority to impose fines on parents if the child continues to be absent from school.

M.3.3.3   Relationship with the community

The teachers agree that the relationship between them and the community is strained and identified this as a key challenge faced. Teachers feel that the community does not support them and fails to show up for meetings.

Even though the relationship between the school and the community is not good, the community still supports the school by helping with the construction of buildings. The head teacher explains that the village council supports the school by enlisting the help of all committee members in securing contributions for the construction of buildings at school

As mentioned above, parents are unhappy when their children are assigned to do domestic scores at teachers' houses during school hours when they should be in class. In addition, parents complain that their children have to walk long distances in order to carry water to school and that the water is of poor quality, frequently making the children ill. They want access to drinkable water for the school and a stop put to their children fetching water for the school and teachers from a well several kilometres away.

> *"They [children] cannot spend a lot of time on digging wells from around 8:00 am to 12:00 noon. They can dig during break time at around 3:00 pm."*

Parents' unhappiness about the loss of instructional time is interesting and a strong indication that they are unhappy about the non-education related activities taking place at school. This view is echoed across all case study sites.

As in all the other schools, corporal punishment is a source of conflict. Parents strongly feel that there should be limits imposed on corporal punishment as children come home injured. Parents also feel that punishment shouldn't lead to the child being unable to continue working at school, because they can no longer hold a pen. In particular, parents reported excessive and unfair use of corporal punishment in a mistaken case of pupil absenteeism which caused anger in the community.

> *"Teachers are beating children on any part of their bodies and they do not limit the number of strokes as legislated. There was a case where a child did not come to school because they were sick and the parent had not reported it. The teacher sent pupils to go and fetch the sick child by force. When the child got to school the teacher beat the child. From that corporal punishment there was a child fainted. Teachers cannot force sick children to come to school."*

M.3.3.4   Summary of selected key findings

Though motivated when starting to work as teachers, teachers at Kishapu Lower report that the multitude of challenges they face has negatively affected their motivation and morale. Teaches believe that they are overworked and underpaid and that their working and living environment is extremely challenging. The lack of training provided by government is viewed as a sign that they have been abandoned and the availability of new textbooks is seen as a problem rather than a solution, as teachers are sceptical that the books cover the correct breadth and depth of the subjects in question. Furthermore, teachers complain that the school does not have enough female teachers and highlight the importance of having female teachers at school, especially at pre-primary level.

The relationship between the school and the community is strained and the question of whether Kisukuma or Kiswahili is spoken is viewed as a power game, unrelated to the language skills of pupils and community members. In addition, teachers reportedly do not respect decisions taken by the head teacher and undermine his authority at school. The community hence views the head teacher as weak whilst the district is seen as unsupportive of the school and failing to deliver on financial commitments made. However, despite all stakeholders agreeing that the relationship between the school and the community is strained, the community continues to support the school in the construction of teacher houses.

M.4 Mpwapwa district, Dodoma

Mpwapwa, Dodoma was selected as one of three qualitative case study districts on the basis of the available quantitative data on literacy, school resources, social and economic resources and the GPI at the lower secondary level. Mpwapwa is a district performing poorly, despite typical levels of school resources and social and economic resources (See Section 11.5). The GPI at lower secondary level in Mpwapwa is reasonably typical amongst intervention districts. Mpwapwa is one of the five districts of the Dodoma Region and most residents live on the arid central plateau, while others live in the mountainous areas where rainfall is higher.

Three schools within Mpwapwa were purposively sampled.

M.4.1 Mpwapwa Better Performing Primary School (Mpwapwa Better)

Mpwapwa Better is one of the largest case study schools (836) and has the most pupils per class (93) and a pupil-teacher ratio that is slightly above average across the case study schools (64). The standard one to three pupil absence rate is the highest of the nine schools (62%) and is higher for girls than for boys. The head teacher is male and has been in post for four years. The school was selected as a comparatively well performing school within the district, with high improvement over recent years. The school is currently almost achieving gender parity.

The community is comprised of agriculturalists and cattle keepers. The agriculturalists cultivate karanga (groundnuts), alizeti (sunflower), ufuta (sesame), rice and mtama (millet in flour form) whilst the cattle keepers keep cows, goats and chickens and are semi-nomadic and frequently migrate in search of better grazing areas taking their children along with them. Cattle keepers are typically polygamous whilst agriculturalists have smaller families. Gypsum mining provides an additional source of employment in the area. Community members belong to the Gogo tribe, speak Kigogo, and are Muslims and Christians. The land is relatively fertile and the government built a well which is now used to support an irrigation system. It was described as a hard-working community, where *"the youths rarely sit around idle."*

M.4.1.1   Teacher capacity, morale and behaviour

Most teachers reported having been inspired to become teachers and having developed a love for teaching. They thought they were joining a profession that would earn them respect and a decent living, where they could impart knowledge to the community and their pupils. However, as in the other case study schools, their expectations were not met. Teachers experience their working environment as extremely challenging and different from what they were told to expect.

> *"I was inspired by my teachers, how they taught us to write, read and to count. They used songs to teach us. They taught us inside and outside the classroom. I really liked how they did this and I said to myself, one day I want to be a teacher and act like them."*

Low salaries are perceived as problematic by all teachers. Teachers feel that they are not respected by the community and think that this is a direct result of their precarious living circumstances. Some community members are wealthier than teachers and teachers think that they look down on them and their problems.

> *"In this community, teachers are not respected, they see a teacher as a broke person with full of debts."*

> *"The community around me, I think their life is easier than mine…..they have time to engage in cultivating while I have to work from 6am to 4:30pm."*

However, other teachers disagreed and felt that the lack of respect was not due to the salaries they earned, but the result of the behaviour and appearance of teachers. In particular, dress code and a professional appearance were mentioned as being related to the level of respect teachers receive from communities. The challenging rural environment and the general state of rural schools was a topic teachers were highly critical of.

> *"Government should prepare a good working environment for a teacher before they send her/him to the school, so that when you report at a school they give you a house rather than you arrive at the school and find there is no place to stay. There are some teachers who report at a school after a month because there is no house for them to go and stay."*

These challenges were perceived as particularly affecting female teachers who frequently request to be transferred to town to be closer to their spouses. Teachers at Mpwapwa Better reported having lost two of their female teachers in recent time.

Training to keep up with frequent syllabus changes was mentioned as an area in which teachers would like to receive support. The relationship between teachers is good and they have a system to help each other when one of the teachers is absent from school. Teachers have also initiated a saving club in order to pool resources as a collective. As these clubs are typically based on trust, this can be seen as a further indicator that teachers at Mpwapwa Better function as a collective.

M.4.1.2 School Leadership

The head teacher at Mpwapwa Better has been in the post for approximately four years. Interestingly, teachers at Mpwapwa Better see parents as the first point of accountability for their attendance at school, whilst the community leader sees the WEC as responsible for ensuring teacher presence at school. Neither sees this as part of the responsibility of the head teacher. Whilst the head teacher is typically viewed as the bridge between the school and the community, this is perceived as the role of the WEC at Mpwapwa Better.

The head teacher highlighted the following challenges at the school: student absenteeism (especially for girls), a lack of desks, malaria for both teachers and pupils, renovation of classrooms, a lack of toilets and no access to electricity, even though the nearby village has electricity. Pupil absence from school is dealt with through the support of the village council which issues fines to parents.

> *"Desks have been a problem because you may find a class with 100 pupils with only 4 desks or three which makes harder for me as teacher to teach with walking around see what my pupils are doing when I am teaching."*

Teachers at Mpwapwa Better discussed a system they have developed whereby teachers sign the attendance register, even if they then leave school. They then ask another teacher to call in on their class so parents are unable to state that a teacher was not present. Therefore, the attendance sheet

does not accurately record the absence of the teachers at school. The SC discussed using part of the financial contributions from the community to provide incentives to teachers with a good record of attendance. The head teacher at Mpwapwa Better is aware of the situation, but feels powerless to change something and crucially doesn't even attempt to get support to address the situation.

> "*I leave them as they are because I don't know what to do… I [do] not even report it to the SC.*"

The head teacher justifies his apathy towards the situation of teacher absenteeism from school by explaining that when he tried to reprimand a teacher, the teacher asked him why he cared given that he received an equally low salary. This is but one anecdote that illustrates that teachers do not feel accountable to the head teacher at Mpwapwa Better.

The head teacher explained that sometimes absence from school, including his own, is due to official trips to the district which can take several days. Teachers complained about not being paid for claims submitted. When the teachers go to the district office to follow up on claims they are ignored and a process that should take a few minutes ends up taking a whole day, infringing on instructional time at school. As a result, teachers are frequently absent from school for entire days, even if the matter they went to raise at the district level should have been resolved in a short period of time. The community leader stated that teacher absenteeism was dealt with by writing a formal reprimand letter to teachers calling them in for a meeting at the head teacher's office where they are asked to justify their absence from school. This system does not seem to work, as parents claim that they frequently ask to see the teacher attendance register when passing through school and that it does not reflect the reality of teacher presence as observed by them. Another adverse behaviour of teachers discussed by a number of stakeholders at Mpwapwa Better was teachers being drunk whilst at work. Whilst all stakeholders perceive the head teacher as having a problem with managing teachers, the head teacher concurs, but points to the fact that he lacks capacity to hold teachers to account.

The WEC at Mpwapwa Better is comparatively active and visits the school once a month. In addition, the head teacher sometimes meets the WEC at pre-arranged locations when the WEC is unable to make it to the school. The head teacher would like the WEC to be provided with an office, transport allowance, and a house close to the school. This would enhance the WEC's ability to fulfil his role. The WEC provides a report to the village council, SC, and the DEO and is also the secretary of several village projects, including for the construction of classrooms and a health centre. The WEC is very involved in the running of the school and actively engages the community. It could be argued that many of the duties he has taken on fall in the realm of the responsibility of the head teacher.

A shortage of financial resources was raised by the head teacher who reports often having to pay for transport to Mpwapwa or Dodoma out of his own pocket. The capitation grant is perceived as insufficient for covering basic school essentials. Typically the head teacher and teachers decide on what ought to be purchased and the SC signs of on the budget which is then sent to the district for approval.

M.4.1.3   Relationship with community

The community leader perceives parents of Mpwapwa Better as being less involved in the school than other parents in the district.

> "*Parents don't help with the raising of funds and other schools have electricity.*"

The relationship between teachers and parents is described as bad and the communication between teachers and parents is often lacking, even though parents discuss raising particularly chronic problems at parent-teacher meetings. Nonetheless, parents also state that if you wait for the SC to

act, you are unlikely to see change happen. Again the low level of teacher salaries is used by teachers as an excuse for a lack of communication with parents.

> *"I cannot follow a parent at home. I am paid low. How can I walk up to their home? I have four years since I became a teacher and I have never visited any parent at their home. I have sent a letter to you and you do not come, I do nothing. I don't know what you have to do to improve education around this area. In town it is easy for the teachers to communicate with parents because they have phones".*

The community is perceived as having diverging views on why education matters and whether it is important. Teachers report that more girls than boys pass exams, but that some parents are unhappy about this.

> *"There was one parent who came to complain why his daughter passed the exams while she was not good in class. He took the matter further to the district education office. He was not happy with his daughter passing exams. Therefore, let`s not talk about a parent coming to complain about not passing more girls."*

It is very interesting that parents feel that it is common for girls to fail their exams, because they have been told that education isn't important for girls. Notably, parents do not really take responsibility for this development.

> *"She [primary school girl pupil] can be changed and think the only thing she wants is to get married so even if you pass someone a salary to talk to her she will not change her mind and she won't pass her exams. All she thinks about is getting married. She is the one to be blamed. No one else".*

Parents clearly do not feel responsible for their daughters' educational choices and early marriages. As a matter of fact, the child is explicitly blamed and held responsible for her behaviour. The head teacher feels that parents have a large role to play and are to blame for the choices girl pupils make.

> *"Parents actually tell their children not to pass because they want girls to get married and boys to keep cattle and agricultural activities."*

Teachers concur and blame pupil absenteeism and low attendance rates on parental attitudes towards education. Teachers claim that children are told to fail the exam by parents (especially fathers) to avoid the cost of sending children to secondary school.

M.4.1.4 Summary of selected key findings

Teachers at Mwampwa Better think that the lack of respect shown to them by community members is the result of the low salaries teachers receive and the conditions they live in. However, the behaviour and appearance of teachers was also perceived as affecting perceptions of the profession. Interestingly, the head teacher also thinks that the lack of respect teachers show him and his office is due to the fact that the head teacher does not automatically receive a higher salary than ordinary teachers.

Teachers at Mpwapwa Better see parents as the first point of accountability for their attendance at school, whilst the community leader sees this as the WEC's responsibility. As a matter of fact, the community leader actually thinks that parents at Mpwapwa Better are less involved than at other schools in the district. Parents clearly do not feel responsible for their daughters' educational choices and early marriages and girls are explicitly blamed and held responsible for their behaviour. The head teacher, however, feels that parents have a large role to play and are to blame for the choices girl pupils make. The head teacher is aware of the problem of teacher absenteeism at the school and mechanisms used by teachers to obscure the scale of the problem, but is unwilling and thinks himself

incapable of addressing the issue. Whilst the head teacher is typically viewed as the bridge between the school and the community, this is perceived as the role of the WEC at Mpwapwa Better. The WEC regularly briefs the village council and chairs village development projects and it could be argued that many of the duties he has taken on fall in the realm of the responsibility of the head teacher.

M.4.2 Mpwapwa Typically Performing Primary School (Mpwapwa Typical)

Mpwapwa Typical is the second smallest school (261 pupils) and has small classes (37) and a low pupil-teacher ratio (29). It has the lowest standard one to three pupil absence rates (7%), which is higher for girls than for boys. The head teacher is female and well established, having been in post for 14 years. The school was selected as a typically performing school with medium to high improvement in recent years. The school is not close to achieving gender parity.

Unlike Mpwapwa Better and Lower, this ward includes the Gogo, Chagga, Bena, Ngoni and Pogoro people groups. Kiswahili is used as the lingua franca in the community and at school due to this diversity. Community members are agriculturalists and cattle keepers, but also entrepreneurs and small business owners that sell crops, local brews and vegetables. Community stakeholders described the standard of living as relatively high and their community as save and relatively healthy. A high prevalence of alcohol abuse was mentioned as a major area of concern.

M.4.2.1   Teacher capacity, behaviour and morale

The teachers at Mwapwa Typical described the best moments of being a teacher as the moment they received their first salary. However, they also describe how this sense of joy was quickly eroded by their responsibilities and the high cost of living. Most of the teachers joined the profession because they were inspired by their own teachers and enjoy sharing knowledge with others.

The teachers are very frustrated with the government and feel that government does not adequately explain what teaching in rural areas entails. The teachers are demotivated and one teacher even suggested that the government reduce the retirement age to 45 from 60 claiming that it isn't possible to do this job for a long period of time. This shows that the teachers want a way out of the system. Teacher report that they feel trapped and have no time to find ways of making extra money.

The teachers listed the following issues as challenges they faced at the school: they have a heavy workload, very low salaries which do not cater to their needs, limited time to make preparations for the many subjects they are expected to teach, a lack of housing for the teachers and too many textbooks that they are told to use when teaching. The community leader feels that the biggest problem with teacher salaries is that they are never paid on time. However, unlike in other schools teachers at Mwapwa Typical feel respected by parents and the wider community.

The SC is aware of the challenges teachers face and is planning to renovate and built new teacher houses as all stakeholders agree that they are in a very poor condition. The parents and the community is contributing to the process of building houses for teachers and even community members who do not have pupils at school support this project. Nonetheless, implementation remains slow, as is also evidenced in the other case study schools.

The frequent curriculum changes are also viewed as problematic by teachers.

> *"There are some top officials who change the curriculum. We teachers are not informed of the changes in the curriculum. We do not participate at all, even though we are the ones who deliver the things that are in the curriculum. For example, we teach English subject while there is no need of it."*

As in many of the other case study schools, teachers at Mpwapwa Typical express a desire to be involved in the process of making curriculum changes.

M.4.2.2 School Leadership

The head teacher feels that the good cooperation between the school, community, village leaders, parents and pupils makes this school different from other schools in the district. In 2013 the school achieved a good result in the national exam and saw 13 of the 19 pupils pass the national exam. Nonetheless, some of the pupils failed to actually attend secondary school for financial reasons.

The WEC has been to visit the school four times this year. He requests an allowance of 10,000tsh per visit, but the school is only able to provide him with 5,000tsh. This is perceived as the reason why he doesn't come and visit more often.

The head teacher displays very good management skills and has adopted clear processes for resolving misunderstandings between teachers and parents. This clarity is also instrumental in the relationship building between all stakeholders involved at the school. When a parent is unhappy they come to see the head teacher who then arranges a meeting between the parent and the class teacher to talk about the problem. If the problem cannot be resolved, it is passed on to the SC for reconciliation. However, so far the head teacher was always able to mediate and resolve conflict at the school.

A challenge the head teacher felt affected her ability to fulfil her dual responsibility of teaching and managing the school was the need to deal with urgent issues during the day. This often required her to leave classes unattended. Nonetheless, the community leader is very happy with the head teacher and describes her as a good teacher who is disciplined, has high ethical standards and values and functions as a role model for the community.

The SC claims that there is no problem with teacher absenteeism at the school. Student absenteeism is reported as existing and teachers feel that the whole community, village leaders and teachers, need to take action if they see a child roaming around during school hours. Ensuring that children attend school is perceived as a collective responsibility. The children that are often absent from school are recorded and their names are given to the SC. The village council then follows up with the families. This is a disciplinary measure which allows the school to exert pressure on parents. Some parents are reported as being displeased at the involvement of the village council. These same parents reportedly also do not follow up on their children's progress at school. The teachers generally express a desire to work with parents on further decreasing the rate of pupil absenteeism from school.

M.4.2.3 Relationship with community

The community leader reports that parents publicly praise teachers for good performance at community meetings at the end of the term in June and December as a way of demonstrating their respect for the teachers' achievements.

Teachers, parents and the SC feel that it is parents', teachers' and government's joint responsibility to ensure that quality education takes place and that all stakeholders have to work together in order to make sure that this is possible. All stakeholders feel that the low pass rate of girls in national exam constitutes a collective failure of all parties involved in supporting the school.

Parents are perceived as very supportive of the school, both in-kind and financially and contribute maize and money to the school. If parents struggle to make the agreed contribution, they are allowed to pay in instalments. If parents are too poor to contribute, they are exempted. Regardless of whether parents contribute or not, all children are provided with food at school. This has caused problems in the community because some parents who do not contribute boasted that their children were allowed

to attend school and received food, whilst parents who were running late on paying their instalments saw their children sent home from school.

The SC contrasts this support provided to the school by parents to the lack of support provided by government.

> *"The government has made us be seen like orphans. But as for the parents, we have enough cooperation because they make contribution in buying of some school facilities or teaching aids. And we have to do so because the government does not give us support".*

The parents see it as their responsibility and the role of a parent to encourage and make sure children go to school. The attitudes towards girl pupils are markedly different at this school and failure at school is not accepted as inevitable, even though teachers also speak about some parents favouring early marriages for their daughters.

> *"I will tell him or her [the student] that student failure is not the end of his or her education. And I will tell the parent to make sure his or her child repeats the class until they pass."*

Parents discussed the problem of poor families sending their daughters to town to work in richer households in order to get additional income for the family. This was perceived as constituting one of the main reasons for drop out of girls from primary school.

M.4.2.4 Summary of selected key findings

At Mpwapwa Typical all stakeholders agree that the relationship between the school, parents, pupils and the community is good. Providing quality education is conceptualized as the responsibility of all actors, necessitating cooperation and support. Subsequently, the low pass rate of girls is viewed as a collective failure. Similarly, the school, SC and village council work together to address pupil absenteeism and community members are expected to report and challenge pupils they see outside school during school hours.

The head teacher functions as a role model within the community and mediates and solves conflict at the school. However, teachers describe themselves as demotivated and unable to find enough time to earn the extra money they need to live. In addition, as in all other case study schools, teachers do not feel that government adequately prepares teachers for the reality of teaching in rural schools. The SC compares the support provided to the school by parents and the community to the lack of support provided by government arguing that government has relinquished responsibility for schools.

M.4.3 Mpwapwa Lower Performing Primary School (Mpwapwa Lower)

Of the nine schools included in the case study, Mpwapwa Lower is the smallest (172 pupils), has the smallest class sizes (25) and the lowest pupil-teacher ratio (17). The standard one to three pupil absence rates are about average compared to other case study schools (41%) and are higher for girls than for boys. The head teacher is male and has been in the post for one year. The school was selected as a school not performing well and with low improvement in recent years. In addition, the school is not close to achieving gender parity and is located at the district boundary.

Community members are agriculturalists and cattle keepers and the majority of parents have only received primary school education. The community is predominately of Christian faith, with a few Muslim families and mostly part of the Gogo people. The community is reliant on rainfall and is prone to famine. A recent drought was reported as having had a devastating impact on the community and the school.

M.4.3.1　　Teacher capacity, morale and behaviour

Most of the teachers reportedly joined the profession because they wanted to share knowledge with what they called under-developed communities and children. Some teachers reported being inspired by their own teachers.

A striking feature at Mpwapwa Lower is that teachers are reported to typically not stay in their posts for more than one year; and head teachers for no more than two years. This is with the exception of a cohort of elder teachers who have been at the school for a long period of time. The morale of teaching staff is incredibly low and is defeatist. The main challenges faced by the teachers at the school seem to be due to the remote location of the school. Teachers complain about the high transport costs incurred when coming to school which have to be covered from their low salaries. They also said it is impossible for them to arrive at the school on time and when they get there they are tired.

> *"You come to school around 8:00 am or 9:00 am and are very tired while the time you leave school is 4:30pm and you spend 3 hours to reach your home place."*

As a result, teachers typically miss the first periods of school and instructional time is severely reduced at Mpwapwa Lower. Teachers acknowledge the problem, but do not perceive it to be their responsibility.

> *"This is a big problem now a days. It is impossible for a teacher to be punctual because we spend a lot of time on coming to school by foot. We miss the first periods and you do not get extra time as a replacement to teach the pupils".*

Only those teachers who live near to the school were reported to arrive on time. Teachers express anger at government allegedly automatically deducting money from their pay check for the teacher houses, regardless of whether there are houses for teachers to live in.

> *"In our salary slip the government deducts 20000Tshs for house rent while we do not stay in those houses. The education sector has a lot of challenges, for sure".*

A lack of teacher housing thus affects both teacher absenteeism and morale at Mpwapwa Lower.

Student absenteeism is also a major problem at this school. Teachers complain that if they teach a section that builds on previous material covered in class, pupils are often unable to follow due to their frequent absence from school. This discourages teachers and affects their morale and is the main explanation provided by teachers for the poor academic performance of children. Teachers also discussed covering each other's classes whenever they have emergencies and receiving financial incentives as a motivation for teaching remedial classes.

M.4.3.2　　School leadership

Mpwapwa Lower is the worst performing school in the standard seven national exams in the district. The main challenges faced by the school according to the head teacher are parental contributions, disagreement over corporal punishment and student absenteeism.

A lack of available transport further affects the head teacher's ability to perform his duties, as the school is situated at the boundary of two districts and it can take up to two days to travel from the school to town. A lot of time is lost and money spent on travelling to withdraw money and collecting information from the district office.

Teacher absenteeism is a big problem at the school and the head teacher feels unable to address this and reports feeling powerless when trying to discipline teachers. He admits that he does not know how to manage a school and complains that he was made a head teacher without receiving any training.

> *"I have been in this position only for one year and never received any in-service training on school management."*

He feels that he is ill equipped to discipline his teachers and that it is unrealistic to expect him to do so. Teacher absenteeism is reportedly an even bigger problem with the elder teachers who have been at the school for a long period of time, as they have established farms near the school and spend large amounts of their time working on them instead of teaching at school. The elder teachers also tell their younger counterparts that the school is bad and that they should transfer as soon as possible.

The school infrastructure is poor and classrooms are few. The school does not have access to water and does not generate additional income from IGAs. The head teacher reports using his own money to buy chalk when the school runs out.

The WEC has been to visit the school four times this year. The head teacher does not think that this is often enough, but attributes the less frequent than desired visits to the fact that WEC does not have transport and hence either has to wait for a lift or receive an allowance from the head teacher before he can visit the school. As in other schools, the WEC is perceived as the true supervisor of teachers who is in charge of monitoring teaching and learning at the school. The confusion around the respective role of the head teacher, the WEC and the chairperson of the SC is illustrated by the community leader who is unable to articulate the responsibilities of the head teacher versus the WEC and sees all school related management issues, including liaising with the community, as residing with the WEC.

M.4.3.3 Relationship with community

All stakeholders claim that corporal punishment is the root cause of the many problems at Mpwapwa Lower and that the continued inability of management to deal with this problem has led to a situation where teachers rarely stay for more than one year and choose to abandon rather than attempt to fix the conflicts and problems at the school.

The SC and parents are unhappy about the application of corporal punishment and report that it can have a negative impact on a child's desire to attend school. However, the head teacher feels that corporal punishment is preferable as it is quick and easy to administer and any other form of punishment requires teachers to invest time to supervise its execution.

> *"Some of the pupils are so young if they are beaten so many times they become afraid of showing up at school. Even sometimes if they have a problem they become afraid of addressing it to the teachers."*

Communication between the teachers, parents, community and pupils is very poor. Most of the pupils cannot speak Kiswahili and especially pre-primary and standard one pupils struggle to communicate with teachers. However, the community wants teachers to only speak Kiswahili in class. Community awareness of education is very low and parents do not provide support to the school. Even the village leadership is unsupportive of the school and fails to create a functioning relationship between the school and the community. A classroom lost its roof during a storm in 2008 and the school and community are still discussing how to fix the classroom and no action has been taken. Furthermore, parents are slow in responding when asked to contribute financially. The head teacher thinks this is partly because people are poor, but even families who can afford to contribute at times don't and he

doesn't know understand why. The school has reacted by only providing food to children whose parents contribute financially.

One reason why the relationship between parents and teachers is poor is the high turnover of teaching staff at Mpwapwa Lower. Parents report that when they do check on their child's progress they find that only three subjects are actually taught at school which leads to further disagreement, as parents had previously been asked to provide seven exercise books to their children. The impact these type of findings have on the perception of the school by the community are very damaging, leading to a further break down in the relationship between the school and the community. Attempts to solve the non-teaching of entire subjects have been made at the end of term meetings and teachers who have failed to work were reported to the WEC.

The community distinguishes between teachers who teach and those who do not. Teachers who are perceived as hard working are respected whilst others are described as lazy and treated with contempt. If communities see teachers outside the school complex during school hours they assume that this means that they are missing unexcused.

The head teacher reports that girls were least likely to remain in school, because parents assume that girls cannot pass national exams.

> *"Parents have never seen a girl pass the national exam."*

In addition, girls are reported to often start working at a young age or are encouraged by their families to marry when they turn 12. The SC claims that as girls reach puberty teachers make sexual advances and that if a girl rejects a teacher, the teacher will react by punishing the girl and giving her bad grades. This reportedly has a detrimental effect on how girls view education and the school and leads to deteriorating performance of girls in the later years of primary school education.

M.4.3.4 Summary of selected key findings

The morale of teachers and the head teacher at Mpwapwa Lower is very low. The head teacher claims that teachers typically do not stay at the school for more than one year, and head teachers for no more than two. A cohort of older teachers has been at the school for a long period of time. The head teacher claims that this group of teachers tells new colleagues to transfer out as soon as possible. Absenteeism levels are reportedly extremely high amongst teachers who pursue jobs and activities on the side and spend large parts of their time on their farms. Furthermore, teachers complained about a lack of housing near the school. Teachers report spending a large part of their salary on transport and claim to be unable to arrive at school on time due to its remote location. As a result, teachers typically miss the first periods of school and instructional time is severely reduced. Moreover, a trip to the district by teachers or the head teacher to tend to administrative issues or collect their salary can take up to two days due to a lack of readily available transport. Whilst the head teacher is aware that he isn't managing the school effectively, he does not know how to address teacher absenteeism and complains that he didn't receive any training when he was made a head teacher.

The relationship with the community is very poor and all stakeholders claim that the conflict over corporal punishment lead to the deterioration of the relationship. The main challenges faced by the school according to the head teacher are a lack of parental contributions, disagreement over corporal punishment and student absenteeism. Parents on the other hand view most teachers as lazy and undeserving of support. Parents report having found out that entire subjects are not being taught at the school. The impact these type of findings have on the perception of the school by the community are very damaging, leading to a further break down in the relationship between the school and the community.

Annex N   Fiscal study concept note summary

The aim of this note is provide a brief overview of the options for costing, level of analysis, data requirements and the next steps in undertaking the exercise.

N.1   The aims of the study

The overall objective of the fiscal affordability study is to estimate the cost and assess the affordability of, scaling up some or all of the activities of the EQUIP-T to other districts after 2018 (OPM Inception report). As part of this, the study will also analyse the within programme cost-efficiency and where possible measure the cost effectiveness of the programme based on the results of the Impact evaluation. In doing so the study aims to answer the following key questions:

1. What were the overall costs of the programme? What were these costs in terms of teachers being trained and schools, districts and regions being reached and supported (cost efficiency)? What are these costs in relation to changes in measured outcomes (cost effectiveness)?

2. What are the likely costs of the programme if expanded to other districts and in other regions? How do these costs differ between a programme implemented in its current operational modality versus a programme that is fully absorbed and implemented by the Government?

3. How do these costs compare to the existing budgets and expenditures of primary education in the targeted districts as well as in aggregate nationally?

4. Is the programme affordable by the Government within the prevailing macroeconomic environment and given the available resource envelope and expenditure commitments?

Following on from the above questions the study is broken down into three components:

a. Analysis of retrospective costs of the EQUIP-T (question 1)

b. Projection of future costs of the programme (question 2)

c. Budgetary and fiscal analysis of future costs (questions 3 and 4)

These are discussed in turn below.

N.2   Analysis of retrospective costs of the EQUIP-T

N.2.1 Cost structure of the programme

The EQUIP-T consists of a Technical Assistance (TA) budget and Programme Support Activities (PSA) budget. The TA budget consist of the salaries of international and national staff implementing the EQUIP-T as well as the operational costs of the national and regional offices. The PSA budget is allocated across five components:

Component 1 - Improving the performance of teachers

Component 2 - Strengthening school leadership and management

Component 3 - District planning and management

Component 4 - Strengthened community participation & demand for better accountability

Component 5 - Strengthened learning and dissemination of results

These components of programme are further broken down into a number of activities and sub-activities, which combined with a budget line on gender and equity as cross cutting issues account for the entire PSA budget.

The budgets allocated to the different components are summarised in Annex table 34. As shown in this table, the three main cost drivers are teacher development (component one), community participation (component four) and cross-cutting gender and equity. These items combined account for close to 70% of total PSA budget.

Annex table 34   **EQUIP-T budget allocations (approximate)**

Programme component	Approximate budget (in £, million)	Share of total PSA budget (%)	Share of total budget (%)
C1: Teacher development	10.0	28	20
C2: School leadership and management	3.5	10	7
C3: District management	5.7	16	12
C4: Community participation	7.3	21	15
C5: Learning and communication	1.7	5	3
Cross-cutting gender/equity	7.0	20	14
Total PSA budget (spent in field of operation)	**35.7 ***	**100**	**71**
Total TA budget (operational costs for CE offices)	**14.3**	**n/a**	**29**
Total PSA + TA	**50**	**n/a**	**100**

Source/Note: * the sum comes to £35.2m, but the PSA budget figure is £35.7m. This costs are taken from EQUIP-T programme inception report.

N.2.2 Analysis of Programme Support Activities

The PSA expenditures highlighted above are captured by activities and sub-activities for each component and reported monthly. This cost structure will be adopted by our costing study and form the basis of analysing programme costs. Under each component there are some costs that are simply transfers of grants at national, regional, district, ward and school level or an award of contracts to service providers at the regional or district level. The breakdown of these costs will be accounted for separately based on templates developed by the Management Agent (MA). These breakdowns should allow us to establish the cost of service provision by district or lower depending on the nature of the grant.

When analysing the PSA activities we aim to answer the follow two questions:

1. Is this activity, something that Government is likely to undertake in future? (Yes, No)

2. If yes, how regularly will this activity be undertaken? (regular monthly, annual, bi-annual, every three or five years)

Answers to these questions will determine the parameters for modelling the future costs of the programme and will be established in consultation with various stakeholders. Recognising the multiplicity of scenarios that could be considered, we will provide a costing spreadsheet that will

enable interested parties to change the parameters of the programme (input sheet) and understand the implications of this on the overall costs.

As highlighted in the previous sub-section 20% of the PSA costs relate to cross cutting gender and equity issues. Currently it is not clear how the costs from this component will be captured. In the event that it is not possible to allocated costs incurred under this item to any activity or component we will apportion these across different components based on size of each component.

In summary the PSA activities will be reported by component (and activities within), disaggregated by district and frequency of recurrence as well as its likelihood of being taken over by the Government.

We will propose the units of analysis for each component (or activity) within the PSA, once the expenditure tracking template for the service providers and other entities receiving grants are established.

N.2.3 Analysis of technical assistance budget

The TA budget will is broken down by Staff salaries (Human resources) and operational costs (e.g. office rental, travel, per diem, etc.). These costs will be input based and captured by region and headquarters. These costs will not be allocated across activities and only reported as overhead costs broken down by the following:

Headquarter Human resources; Operations (Office rental; Travel; Per diem; Office Stationary and equipment)

Regional Human resources**;** Operations (Office rental; Travel; Per diem; Office Stationary and equipment)

The costs structures at the regional level as based on a very similar operational modality across region and based on a similar staffing structure. This should provide sufficient information on the likely cost of monitoring and support structures under programme expansion based on the existing project based operational modality that is unlikely to be replicated by the Government. It is important to note that if there are costs that the TA is providing and which will have to be taken over by the Government that this will need to be captured. However our understanding is that this is not the case.

N.3 Future costs of the programme

N.3.1 Parameters of interest

In analysing the existing costs of the programme, a number of parameters (e.g. cost of INSET training per teacher, grant per ward or district, cost of policy formulation, etc.) will be obtained and used for projecting the cost of the programme if scaled up. However in addition to this the model will require additional demographic as well as education statistics to support the projection process. Information that is likely to be required include district enrolment rate, population projections for number of children of primary school age, number of teachers by district, number of schools and the number of Ward Education Coordinators by district. Availability of such statistics is currently being investigated.

N.3.2 Modelling costs

The costing model will look at two models of operations:

I.   Cost of expanded programme under existing operational modality

II.   Cost of expanded programme if fully absorbed by the Government

Under the first model we use existing cost structure (i.e. one office per region) for expansion, taking into account some activities that may not be replicated, at least in the short to medium term (e.g. teacher performance framework). For the second model we need to establish which activities are likely to be undertaken by the Government and determine their frequency.

The future cost of the programme will be projected under a number of scenarios that in addition to the above modality, will include the pace of expansion and target date by which the programme reaches the entire districts within the country. These parameters will be determined in consultation with the Government, DFID and MA. The future projections will be developed within an excel sheet that can be used as a costing tool for planning and budgeting purposes in the future.

N.4   Analysis of costs within the budgetary and fiscal process

N.4.1 Costs relative to existing education budgets and expenditures

Once the costs of the programme are determined, it is useful to place these costs in the context of existing budgets and expenditure of the education sector, with particular focus on primary education at the district. The study will aim to assess the costs of the programme in relation to the detailed budget of primary education in the piloted districts. To do so the team will require budgetary data for the current and past budgets as well as actual disbursements and expenditures for Personal Emoluments (Budget submission Form 8 b), the development budget (including Capital development grant) as well as the other charges budget. The District Education Officer and District Planning Officer should have access to these files and information. Since the third component of the programme is meant to support the planning and budgeting process at the district level we expect that this information will be collected by MA and shared with OPM. A note with the detailed data requirements and form will be shared with MA and DFID.

In principle this information should also be available at the national level through the Prime Minister's office regional administration and local government (PMORALG) in Dodoma and steps will be taken to obtain these, although support from DFID and MA will be required.

This process will capture all information that is on budget. There may be some projects and sources of finance that remain off budget, we aim to better understand the extent to which this is the case through our team's discussions with PMORALG, DFID as well as MA based on work they are carrying out related to public financial management.

N.4.2 Costs in the context of macroeconomic environment

Fiscal affordability of a national programme will depend on the macroeconomic environment of the country and the resource envelope available against existing expenditure commitments. This study will provide a summary of the macroeconomic environment and the trajectory of key macroeconomic indicators based on existing analysis and projections notably IMF Article IV mission reports, Public Expenditure Reviews conducted as well as publically available documents and data from the Government of Tanzania. The costs of the national programme will be analysed within this context.

N.5   The next steps

Once the MA provides the detailed financial reporting templates for service providers and grant making, and the level detailed captured and reported on regularly is established, we will develop a template of the costing analysis and share with DFID and MA. Once agreed this will form the basis of

our analysis. Following this we will discuss and agree a protocol for regular data sharing and update with MA.

Annex O   Impact evaluation governance

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 the Evaluation Director. The Project Manager is the Secretary for the OPM EQUIP-T Governance Team, which will meet twice a year (virtually) and more often on a needs basis.

Management is executed by the Project Manager, an OPM staff member, 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 survey team 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 team and a technical advisory team. The Project Manager and the Evaluation Director together engage with stakeholders including the MoEVT, PMO-RALG, government education agencies (for instance, NECTA and TIE), national researchers, and media. They also work with the MA on information sharing for the IE.

Quality Assurance is provided internally using a two stage process. In the first stage, each key activity and output is reviewed by the relevant technical advisor: Patrick Ward for evaluation, Juan Munoz for sampling, Ingo Outes Leon for quantitative design, Ian MacAuslan for qualitative research design and Georgina Rawle for the fiscal affordability study. In the second stage, there are three reviewers of each output generated by the IE: Paud Murphy, Evaluation Director, Professor Herme Mosha, Lead National Education Specialist, and a leading academic researcher in the field of education and economics (Dr Caine Rolleston for the baseline outputs). 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 P   Specialist Evaluation and Quality Assurance Services matrix

Area		Question	Addressed in	Comment
1	1.1	Is the product accessible to the intended audience (e.g. free of jargon but using technical language clearly, written in plain English, logical use of chapters, appropriate use of tables, graphs and diagrams)?	-	
	1.2	Is it clear who has carried out the work? Are all primary and secondary sources clearly referenced (internally or externally) or explained?	References sections (Vol I and II)	
	1.3	Is an executive summary included, and can it stand alone as an accurate summary of the main product? Does it summarise the key decisions and choices made?	Executive Summary section (Vol I).	
	1.4	Does the report contain a clear explanation of the phases of the baseline design and operation including how methods were chosen and who chose them, data collection and preparations, analysis and reporting outlined.	Part F Methods (Vol II). Annex E: Quantitative data collection, cleaning and analysis (Vol II), Annex J Details of qualitative sampling, tools and fieldwork (Vol II), Annex B Annex B Stakeholder consultations (Vol II).	
	1.5	Do the annexes contain – at the least – the original TORs, the evaluation framework (including evaluation questions), and a bibliography?	The TOR and the evaluation framework are presented in Annex A (Vol I). Bibliography is presented in the References sections (Vol I and II).	
	1.6	Do annexes increase the usefulness of the product?	-	
	1.7	Is source data included and referenced as appropriate? Are tables and other visual display of data clear and with purpose?	-	
	1.8	Have any departures from the original TOR been adequately explained and justified?	Annex A (Vol I) describes in detail how and why TOR evolved away from the original version.	
2	2.1	Does the product provide a sufficient description of the intervention to be evaluated and formalise this with reference to baseline creation? 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. There should be a clear line of sight from this explanation through to the choices made on the baseline.	Section 1.2 1.2 Overview of the EQUIP-T programme and Annex D (Vol I), Annex B Impact evaluation districts (Vol I), Section 3.1.2 (Vol II) Identifying the control group.	

2.2	Is the process by which the choices about the baseline were made clearly explained? Have key stakeholders been identified and involved, for example in choosing weightings for composite indicators or in selecting reasonable counterfactuals?	Annex B (Vol II) explains how and when stakeholders were involved. Section 3.1.1 (Vol II) Assessing the impact of the EQUIP-T programme as a whole and Section A.2, Annex A (Vol I) explains how choices about the evaluation were made.	
2.3	Does the product provide a relevant and sufficient description of whether and how contextual factors around the evaluation (local, national and/or international) and practical issues have influenced the baseline design?	Annex B (Vol II) stakeholder consultations, Section 3.3 (Vol II) Instrument development, section 2 Pupil learning outcomes (Vol II) based on Nigeria teacher development needs assessment (TDNA) and pupil learning assessment based on BRN EGRA and EGMA and lesson observations based on UNICEF teacher study in Tanzania, section 3.2.2 (Vol II) Stage 1: Selection of control districts.	
2.4	Does the product identify key linkages between the intervention and other relevant projects / programmes / donors, particularly with respect to counterfactuals and explaining attribution? If no linkages are identified, does the product justify why other projects / programmes / donors will not be relevant to the interpretation of changes compared to the baseline?	Section 1.3.3 (Vol I) and section 3.2 (Vol II).	
2.5	Does the document clearly show how the evaluation question and other factors lead to the strategy for baselining and to a null-hypothesis/counterfactual? Are the proposed baseline variables clearly explained as dependent or independent variables (y/x variables)?	Section 1.3.1 (Vol I) 1.3.1 Overall impact evaluation, section 3.4 (Vol II) Impact indicators, Annex Table 1, Section A.2, Annex A (Vol I) Original and agreed terms of reference, section 3.1 (Vol II) 3.1.2 Identifying the control group, Annex K (Vol II) Qualitative evaluation matrix, Annex L (Vol II) Key qualitative questions and data sources.	
2.6	Does the product describe how the baseline addresses the evaluation's accountability and/or learning purposes? Are the decisions on the baseline assessed against these and cost/benefit of data and analysis?	Purposes of evaluation are described in Section A.2, Annex A (Vol I)	The impact evaluation will be an accountability and learning evaluation with relatively more focus on the former.
2.7	Does the product justify the timing of the baseline and the forward data collection for comparison? Is there sufficient discussion of comparison periods? Has backwards extension of the baseline for longer comparison been considered? How far forward will data be collected to show post-project effects?		The timing of the baseline (2014), midline (2016) and endline (2018) were determined by DFID to allow results to feed into the policymaking cycle in Tanzania. Quantative and qualitative research are adjusted to work together (Section 2.1, Vol II)
2.8	Does the product clearly show the range of alternative choices that were considered for the baseline and the criteria by which the final decision was made during the inception phase?	Section A.2, Annex A (Vol I) explains how choices about the evaluation were made.	

3	3.1	Is the proposed baselining methodology described and justified in appropriate detail?	Part F (Vol II) describes the methodology.	
	3.2	Are these methods broadly appropriate for addressing the evaluation questions? Is the explanation why they have been chosen complete and clear?	Section 3.7 (Vol II) and 4.8 (Vol II) address limitations to the methodology. The rest of Part F describes the methodology in detail.	
	3.3	Are primary and secondary data sources appropriate, adequate and reliable? Has a balance been struck between cost and benefit of primary and secondary data?		The IE collects primary data as it would not be possible to measure programme impact on pupil learning outcomes and teacher professional capacity using secondary sources as not available or fails to capture the core areas the programme is seeking to impact.
	3.4	Is the overall sampling and stratifications strategy adequately described and appropriate? Will the methodology enable the collection and analysis of disaggregated data to show difference between groups through appropriate stratification and sample sizes?	Section 3.2. (Vol II) describes sampling for the quantitative part. Section 4.4. (Vol II) describes sampling for the qualitative part.	
	3.5	Is the mix of qualitative and quantitative data used appropriate? If focused on one or the other, is it adequately explained why?	Mixed methods approach is explained in section 2 (Vol II).	
	3.6	Is the accuracy and robustness of source data and analysis adequately addressed?		See Annex E, (Vol II) Quantitative data collection, cleaning and analysis and Annex J (Vol II) Details of qualitative sampling, tools and fieldwork.
	3.7	Does the baseline offer multiple approaches to confirm assertions and/or triangulation of data? If not, is there a clear rationale for doing otherwise? Is there unnecessary repetition of the same data purporting to be confirmation?	Part B (Vol I) compares IE baseline results to results from other source when possible, e.g., the BRN 3Rs EGMA/EGRA for Tanzania and the World Bank's SDI survey for Tanzania. Mixed methods approach in section 2 (Vol II) describes how findings/information from qualitative and quantitative approaches will inform each other. The qualitative research follows ups on selected quantitative baseline results to identify areas of agreement/contradiction.	

3.8	Has there been realistic consideration and explanation of data availability both for the baseline and future comparisons (over time, space) including the possibility of definition change in sources between the baseline and follow up data? What is the stability of the sources of the underlying data (may the series stop, is it a panel)? What may happen if data is unavailable later in the evaluation?		Midline and endline data will come from surveys by OPM using the same instruments and also the same fieldwork teams to the greatest extent possible. The same 200 schools surveyed at baseline will be surveyed again in 2014 and 2016 (the pupils and teachers are not panelled). The likelihood of some of the schools closing down over the four year period is small.
3.9	Is there fully convincing explanation of data collection and cleansing approaches? Is it clear that unique identifiers have been used? Or if not explanations of matching / fuzzy matching and limitations around this?	Annex E (Vol II) Quantitative data collection, cleaning and analysis and Annex J (Vol II) Details of qualitative sampling, tools and fieldwork.	
3.10	What are the assumptions made about further generalisability of results? How do the assessments of bias affect these assumptions? Is there over-claiming on the data and analytical implications?	Section 1.3.4 (Vol I).	
3.11	Is there a reasonable treatment of statistical significance? In cases where results are not significant are conclusions still implied, how are they explained?	Section 1.4.2 (Vol I).	
3.12	Are the assumptions around causality, counterfactuals, and conclusions soundly drawn? Are these reasonable and fully explained?	Section 1.4.2 (Vol I) explains difference of statistical significance and causality; section 3.5 (Vol II) explains how impact estimation will be done; section 3.1 (Vol II) describes counterfactual issues.	
3.13	Is the baseline designed to be useful for assessing the cross-cutting issues of gender, poverty, human rights, HIV/AIDS, environment, anti-corruption, capacity building, and power relations? Where this is not so, are the practical / cost reasons clearly explained?	The impact evaluation will address aspects of gender and poverty through quantitative and qualitative methods. Quantitative results are disaggregated by gender and poverty status in Part B Results (Vol I).	The EQUIP-T programme does not (directly) target human rights, HIV/AIDS, environment, anti-corruption and power relations and therefore the impact evaluation will not cover these areas. The Technical Proposal proposed political economy analysis to be part of the impact evaluation but DFID did not select this option.
3.14	Does the baseline allow exploration of Paris Declaration principles within the context of this intervention?		This is not relevant for the impact evaluation.
3.15	Are any other methodological limitations acknowledged and their impact on baseline quality discussed? Are the limitations acceptable and/or are they adequately addressed?	Section 3.7 (Vol II) 3.7 Possible risks and limitations of quantitative component and 4.8 (Vol II) Possible limitations of the qualitative component.	

4	4.1	Does the methodology respect concerns around gender, age, ethnicity, caste, religion, geographic location, ability, socio-economic status and hard to reach groups? If not, why not?	Annex C (Vol II) Permits, consent, confidentiality and datasets.	
	4.2	Were the practical arrangements around collecting information form vulnerable individuals or groups adequately considered and not expected to affect the data quality?	Annex C (Vol II) Permits, consent, confidentiality and datasets.	
	4.3	Does the baseline allow for consideration of DFID's commitment to human rights based approaches? If not, why not?	-	
	4.4	Do the governance structures underlying the decisions on choice of baseline include diverse perspectives, and are such perspectives free of control from organisational influence and political pressure?	Annex B (Vol II) explains how stakeholders were consulted. Annex D (Vol II) explains how dissemination to different stakeholders will be implemented. Annex C (Vol II) explains how ethical consent was reached from different stakeholders. Annex 0 describes the IE governance structure.	
	4.5	Are all data collection instruments (and those of secondary data) clearly shown to not have any ethical problems associated with them, or where there are questions are they adequately addressed?	Annex C (Vol II) explains how fieldwork followed ethical standards. Section 3.3 (Vol II) explains how instruments were developed.	
5	5.1	Where a mix of existing or new data are used is this appropriate? Are the sources interoperable?		The baseline uses primary data for the analysis. Secondary data was used for sampling (section 3.2 Sampling strategy and sample size, Vol II) and matching purposes (section 3.5 Impact estimation, Vol II).
	5.2	Is the unit of observation appropriate? Does a different unit need to be used for analysis and collection? Is this explained and allowed for?	Different units of observations are used, depending on the indicators of interest. This is explained in Section 3 (Vol II).	
	5.3	Have suitable methodologies for attribution been considered?	Section 3.5 (Vol II) Impact estimation.	
	5.4	Is the selection of comparison groups or areas adequately explained and justified?	Section 3.1.2 (Vol II) Identifying the control group.	
	5.5	Has the issue of bias been addressed? Including geographical, gender, other group biases? Interviewer or language bias?	Sections 3.1.2 (Vol II) and 3.2 (Vol II) explain how bias during sampling has been dealt with. Section 3.5 (Vol II) explains how bias in terms of impact estimation will be dealt with. Annex E (Vol II) explains how during data collection, bias was reduced via training, piloting, and quality assurance.	

5.6	Has the issue of non-response (and associated bias) addressed?	Unit response rates are presented in section 3.2.3 (Vol II), Intended and actual number of observations are shown in Annex H Detailed tables for Part B results (Vol II). TDNA results adjusted for item non-response are shown in H.6 Supplementary teacher subject knowledge analysis (Vol II).	
5.7	Does the methodology use compatible territorial comparisons?	Section 3.2 (Vol II) explains how sampling was implemented, including selection of control districts.	
5.8	Is the sampling done on a sound basis? Is the sampling regime linked properly to the question being addressed (for example is the difference in sampling for a proportion estimate or an estimate of a level of a variable, or comparison between variables properly explained and justified?)	Section 3.2. (Vol II) Sampling strategy for quantitative and section 4.4 (Vol II) Sampling for qualitative.	
5.9	Was the sampling probability or non-probability? How does the analysis take this into account?	Section 3.2. (Vol II) Sampling strategy for quantitative and section 4.4 (Vol II) Sampling for qualitative.	
5.10	Is the baseline intended to inform a change assessment or is it intended to provide an accurate estimate of the current level of variables? Is this appropriately dealt with?		The baseline is intended to provide an accurate estimate of the current level of variables.
5.11	Is the training/recruitment/use of interviewers and other data collectors clear and any limitations around training or ability explained and accounted for?	Annex E (Vol II) Quantitative data collection, cleaning and analysis. Annex section J.3 (Vol II) Details of qualitative sampling, tools and fieldwork.	
5.12	Is there an explanation of the field testing/ translation / cognitive testing of questionnaires and other techniques explained and issues encountered properly described and mitigations explained?	Section 3.3. (Vol II) Survey instruments.	
5.13	Is an explanation of post-data collection validity checking explained and any unusual results described, and discussed with any mitigations?	Annex section E.7 (Vol II) Data validation, cleaning and analysis for quantitative and section J3 (Vol II) Fieldwork for qualitative.	
5.14	What thought and explanation of effects on results (and biases to results) was given to the reward structure to, and the costs on, respondents of data collection?		There are no rewards (for participating in the survey or for changed behaviour) for respondents.
5.15	What were the supervisory arrangements around the interviewing / data collection – were there any risks to data integrity, what mitigations were in place?	Annex sections E.5 and E.6 (Vol II) for quantiative and Annex sections J.3.2 Daily debriefs and team checks in the field and J.3.3 Data treatment and collection for qualitative.	

	5.16	Where surveys or other data collection were carried out by a third party organisation is there sufficiently convincing governance and management in place to be sure of data quality?		OPM conducted all the qualitative and quantitative fieldwork.
6	6.1	Have the potential users and stakeholders, and the ways in which the evaluation could be used, been identified?	Annex B (Vol II) explains how stakeholders were consulted. Annex D (Vol II) explains how dissemination to different stakeholders will be implemented.	
	6.2	Is there any evidence that the priority questions and issues of key users and stakeholders have been addressed in the baseline evaluation?	Annex B (Vol II) Stakeholder consulations.	
	6.3	Does the design of the baseline, analysis and presentation of data and analysis and plans for future data use enable a transparent process that engages and meets the needs of all users, including primary stakeholders?	Annex D (Vol II) explains how dissemination to different stakeholders will be implemented and Annex C (Vol II) Permits, consent, confidentiality and datasets.	
	6.4	Are stakeholders affected by the intervention properly considered in terms of their data protection and access to their own personal data?	Annex C (Vol II) Permits, consent, confidentiality and datasets.	
	6.5	Is there clarity around the final ownership / copyright of data, analysis and reports? Are the arrangements for storage and accessibility of any data generated through the work clear?	Annex C (Vol II) Permits, consent, confidentiality and datasets.	
	6.6	Are there suitable methods for communication appropriate to meet the diverse needs of stakeholders, including gender concerns, and access for marginalised or non-literate groups affected by the intervention?	Annex D (Vol II) explains how dissemination to different stakeholders will be implemented.	

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 impacts 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 DIFD, 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.