

EQUIP-Tanzania Impact Evaluation

Endline Quantitative Technical Report, Volume I

Results and Discussion



FINAL REPORT

Georgina Rawle, Michele Binci, Gunilla Pettersson Gelande, Jana Harb, Paul Jasper, Safa Khan, Deo Medardi, Alessio Romarri, Michelle Rorich, and Nicola Ruddle.

25 January 2019

Acknowledgements

The authors would like to thank all the individuals who have contributed to the Education Quality Improvement Programme in Tanzania (EQUIP-T) impact evaluation to date, and to producing this report. These include:

- Former and current members of the evaluation's Reference Group, who are overseeing the evaluation, and have provided valuable technical advice on all of the reports produced so far. Current Reference Group members are detailed in Annex B.2 in Volume II.
- EQUIP-T managing agent (MA) staff: former and current staff from the MA have been generous with their time, and have shared documents and data, as well as answering numerous questions about the programme.
- DFID advisers: former and current education advisers, and results advisers, have provided insightful feedback and guidance at key stages of the evaluation.
- Coordinators of other national education programmes: advisers from LANES and Tusome Pamoja have taken time to share documents and to explain programme activities to the evaluation team.
- All the individuals who contributed to the baseline and midline evaluations: those who provided technical inputs to the design, the field teams who collected the data, and the analysts. Full acknowledgements can be found in the baseline and midline evaluation reports (Oxford Policy Management (OPM), 2015a and 2017a).
- The endline quantitative fieldwork team from OPM Tanzania, managed by Ignatus Jacob, Deo Medardi, Jana Harb, Andreas Kutka and Diego Shirima. The fieldwork was successfully completed thanks to the hard work of the enumerators and their supervisors, who visited 200 primary schools in 25 districts across Tanzania. Fieldwork team members are listed in Annex D in Volume II.
- Last, but not least, all the survey respondents: head teachers, teachers, pupils and their parents, who generously gave their time and shared information and views.

This report was reviewed by Professor Herme Masha (USDN), Paud Murphy (independent consultant, and former World Bank Senior Education Specialist for Tanzania), and Dr. Caine Rolleston (University College London, Institute of Education). The pupil learning analysis was reviewed by Dr. Joshua McGrane (Rasch measurement specialist, University of Oxford).

All opinions expressed, and any mistakes, remain the responsibility of the authors.

Executive summary

The Education Quality Improvement Programme in Tanzania (EQUIP-T) is a six-year (2014–20) Government of Tanzania programme with a £90 million budget funded by the UK Department for International Development (DFID). The aim of the programme is to increase the quality of primary education and to improve pupil learning outcomes, in particular for girls. Initially, the programme was intended to run for four years, with activities targeted at five, and later seven, of the most educationally disadvantaged regions in Tanzania.¹ In 2017 the programme was extended for a further two years, and the extension introduced some new sub-components to the seven regions, and introduced a reduced package of interventions to two new regions.

The programme consists of five broad components: (1) improved access to quality education; (2) strengthened school leadership and management (SLM); (3) strengthened district planning and management; (4) stronger community participation and demand for accountability in education; and (5) improved learning and dissemination. In total, there are 10 sub-components within these components.

This report presents the findings from the first part of the endline round of the impact evaluation of EQUIP-T carried out by Oxford Policy Management (OPM). This follows two previous rounds of research: a baseline in 2014 and a midline in 2016. The main aims of this first part of the endline evaluation are to estimate the impact of EQUIP-T on pupil learning achievement, and to assess the effectiveness of the school- and community-level EQUIP-T interventions, after nearly four years of implementation (44 months). This part of the impact evaluation focuses on the sub-component to improve teacher performance (1A), and covers aspects that can be measured quantitatively under SLM capacity building (2A), the school information system (SIS) (2B), community participation and accountability (4A), and providing a conducive learning environment for marginalised children, particularly girls and children with disabilities (4B).

The results of this first part of the endline impact evaluation are intended to inform further adjustments to the programme before it finishes in January 2020, as well as to promote accountability and lesson learning for DFID and the Government of Tanzania. This is based on the quantitative assessment of impact and effectiveness. The findings will also help to guide the design of the second part of the endline, which will include qualitative research and a cost study.

The impact evaluation has a mixed methods design based on three rounds of quantitative and qualitative data collection, supplemented by secondary data, including financial data from the programme. The main data source for this part of the endline is three rounds of a quantitative survey from a panel of 200 primary schools (100 EQUIP-T treatment schools, and 100 control schools), which started with baseline data collection in 2014. The evaluation uses a quasi-experimental approach to quantitative estimation of impact that combines propensity score matching (PSM) with difference-in-differences (DID). The survey also allows trend analysis in relation to the programme treatment areas, in order to illuminate reasons for change or lack of change.

Key findings on pupil learning

This impact evaluation finds that the programme has had a positive impact on both literacy and numeracy skills for pupils in treatment schools, compared with pupils in control schools. The positive impact on Kiswahili skills is evident in the fact that more Standard 3 pupils are in the top achievement band, and fewer in the bottom band, than would have happened in the absence of

¹ There are 26 regions in mainland Tanzania.

EQUIP-T. Between baseline and midline the programme brought pupils out of the bottom band, and between midline and endline the programme moved more children into the top band, for Kiswahili. In maths, EQUIP-T had an impact between midline and endline in maintaining maths scores: in the absence of the programme, maths performance would have dropped. These results clearly show that EQUIP-T has had some success in achieving its overall impact-level goal.

Improvements in girls' learning have been more pronounced than for boys in schools in the programme, with an attainment gap opening up in Kiswahili (with girls performing better than boys), and the boys' lead in maths narrowing. There is some evidence that EQUIP-T has contributed to larger gains for girls: teacher interactions with pupils became more gender-balanced, though this improvement was seen between baseline and midline even though specific gender response pedagogy training was rolled out after midline.

There has also been significant progress since baseline in narrowing the learning gap in literacy skills for children who do not speak Kiswahili (as a main language) at home. However, pupils from households who speak Kiswahili at home still perform far better than those who do not in both Kiswahili and maths, even after taking poverty into account. Since baseline, this impact evaluation has stressed that the overwhelming majority of pupils do not speak Kiswahili as their main language at home. As pointed out at midline, it is imperative for teacher training to take account of this barrier to learning, and to support teachers with strategies to overcome it, in order to ensure the training is suitable for the context. Whilst it is positive to see that the general teaching practices covered in the in-service training have supported a narrowing of the learning gap in literacy skills, this raises the question of whether the programme should have been more tailored to address this issue.

Explaining EQUIP-T's impact

EQUIP-T's positive impact on pupil learning is highly commendable and important, and it follows closely the interventions on early grade teacher training: the in-service training focused on Kiswahili in the first two years, and there was a positive impact on Kiswahili performance; the focus shifted to numeracy in the second two years, and then impact on maths performance was achieved. **Given that participation in the teacher in-service training is relatively high, and it has been broadly implemented as intended, it is likely that this aspect of the programme has made the biggest positive contribution to impact.**

Furthermore, the generally low levels of pupil learning outcomes relative to curricula expectations highlight the relevance of EQUIP-T; its focus on in-service training for teachers and changing learning outcomes for children has been highly appropriate. **However, when unpicking many of the intermediate outcomes expected to contribute to pupil learning, according to the programme theory of change, the endline finds that the programme has not been successful in achieving all the expected effects. This raises the question of whether the potential impact on pupil learning could have been substantially greater.** Alternatively, assumptions about the changes required to achieve certain intermediate outcomes may have been misguided.

Meanwhile, external influences are likely to have put negative pressure on the programme. The introduction of fee-free basic education led to a surge in enrolment, and the Government's professionalisation policy caused high recent turnover of head teachers and WEOs. The programme's managing agent (MA) also pointed to delays in Government approval of materials and fund release for local government authorities (LGAs) as barriers to timely implementation. The achievement of impact despite these factors is noteworthy.

Teachers' performance

Summary of implementation under Component 1A: teacher performance

- In-service training targeted at all Standards 1 and 2 teachers and sometimes including other teachers, broadly in order: 13 modules of training on early grade Kiswahili literacy; training on 3Rs curriculum and syllabus; nine modules of training on early grade numeracy; and training on gender-responsive pedagogy. Training was a mix of residential and school-based.
- Provision of teaching and learning materials (TLMs) for the lower standards, including 'big books', supplementary readers, teacher 'read-aloud' books, and literacy and numeracy teaching aid toolkits.
- Posters on positive classroom behaviour management and a positive learning environment.
- Introduction of a communities of learning (COL) concept, using school-based in-service training, weekly school performance management meetings (SPMMs), and quarterly ward cluster reflection meetings.

At endline, almost all targeted teachers have attended the in-service training, although the coverage of modules has not been complete. The incomplete coverage of modules relates in part to a context of very high teacher turnover: just over half of early grade teachers from baseline are still in the schools at endline although most moved within EQUIP-T districts. With such high turnover, as reported in the baseline and midline phases of the impact evaluation, and movement of teaching responsibilities between lower and upper grades, it is not surprising that many teachers have not attended training on all the modules.

Schools' implementation of the school-based in-service training model has been improving since 2015; however, there is still wide variation in the number of sessions held and teachers' attendance. High turnover of teachers and in-service training coordinators (INCOs) is particularly problematic for this aspect of the training model. Challenges to attending school-based sessions reported by teachers include the lack of allowance, low motivation and inconvenient scheduling. Meanwhile, teachers want to see more training materials provided. These findings suggest a risk to the sustainability of the training model – although it is designed to be low-cost and locally-driven, capacity to maintain implementation appears to be low at the school level.

The endline finds mixed results on the use of positive teaching practices. Teachers' interaction with pupils across the classroom, a measure of inclusion of all pupils in the lesson, improved substantially between baseline and midline. However, this has not improved further since midline and in fact teachers are interacting less with pupils at the back of the room – likely a reflection of the larger class sizes. Similarly, the gender balance of teachers' interactions with pupils improved markedly between baseline and midline and this was sustained to endline, despite the sharp rise in classroom overcrowding. However, while teachers are demonstrating some positive behaviours more, use of many of the positive teaching practices measured by the survey has significantly worsened. Thus on the pedagogical practices measured by this impact evaluation it is unclear if EQUIP-T has helped to improve teaching quality overall. Teacher training needs to be tailored to ensure that teachers are able to deal with large pupil numbers.

Teachers are using different instructional materials more frequently in their lessons. However, despite TLMs being distributed to schools, EQUIP-T's materials are typically not found in classrooms and are not being used. This raises the question of whether the materials are relevant for teachers in these schools, since they are not choosing to use them, and whether the in-service training could have done more to support teachers in knowing how and why to use the materials.

Overall, there has been a reduction in teachers' absenteeism from the classroom since baseline, for Standards 1 and 2 teachers, and most of this change happened before midline. There is weak evidence that EQUIP-T had an attributable impact on classroom absence for teachers overall, and

absenteeism levels are still extremely high. Teacher attendance in the classroom is a fundamental assumption underlying the programme theory of change, which raises a question as to whether the programme could have identified better ways to address this than its indirect route of improving motivation.

As mentioned, pupils' Kiswahili performance improved most strongly between baseline and midline, reflecting the focus of the in-service training. The slowing of improvement after midline coincides with a more challenging context for teachers: class sizes have grown substantially with the introduction of fee-free education, the range of ages has increased (increasing the proportion of over-age pupils), and more children now come from poorer households and are occupied with more work outside the home.² At the same time, the maths results show that whilst these challenging contextual factors may have put a downward pressure on maths performance, the EQUIP-T focus on numeracy in the in-service training after midline played a part in mitigating this.

SLM

Summary of implementation under Component 2A and 2B: SLM capacity building and SIS

- Training for head teachers, assistant head teachers, and Ward Education Officers (WEOs) on: module 1 – school quality standards and national school leadership competency frameworks, and roles of head teachers; module 2 – leading EQUIP-T school-level initiatives, including parent–teacher partnerships (PTPs), SIS introduction; and module 3 – school development plans (SDPs).
- Training for head teachers and WEOs on holding SPMMs.
- Provision of tablets to all head teachers, accompanied by initial training on the SIS.

The programme has not been as effective as anticipated at rolling out SLM training: around three-quarters of head teachers have attended SLM training in the last two years, and those who have said the training was helpful in regard to knowing their responsibilities and in relation to teacher management. Again, the incomplete coverage is linked to high turnover, with only a quarter of the head teachers from baseline still in the school at endline. High turnover was highlighted in the midline evaluation, and this raises the question of whether the programme could have done more to mitigate the challenges of high turnover, in order to be a relevant and sustainable training model in this context.

At midline, SDPs were more available than at baseline and were felt to improve the relationship between head teachers, teachers, and communities.³ Unfortunately, the availability of SDPs has dropped since midline, coinciding with the tailing off of SDP training and high head teacher turnover.

There have been some positive improvements in head teacher leadership and management: head teachers are managing teachers' performance through checking more lesson plans; holding more staff meetings; and they are giving more rewards for teachers. On the other hand, the frequency of lesson observations by head teachers has fallen significantly since midline.

EQUIP-T's SIS has so far not shown signs of effectiveness. Almost all schools have tablets; however, less than half of head teachers have had SIS training (again, likely a result of turnover), only a small proportion of schools have up-to-date enrolment and teacher data entered on the tablet, and even fewer are entering any attendance data. There are many problems: it takes too long to enter the data, most schools do not have functional electricity for charging the tablet,⁴ and internet connectivity is

² Whether the shift in pupil characteristics was due to a national change, or a rebalancing of the make-up due to the enrolment of pupils who previously would not have enrolled, is unknown.

³ Evidence from the qualitative research at midline.

⁴ This was a reported problem by head teachers even though the solar chargers are supposed to be provided with the tablets.

poor. Given these challenges it is perhaps unsurprising that so few schools are using the SIS as a tool for school management. A more substantial assessment of needs and practicalities for entering and using the data during development might have helped to ensure this intervention was more relevant and thus more useful.

Community participation

Summary of implementation under Component 4A: community participation and accountability

- Training for school committees (SCs) on roles, responsibilities, processes, school improvement, PTP formation, and grant management. Training for PTPs on their role vis-à-vis SCs.
- Provision of PTP grant 1 for PTP activities and general school improvement, and PTP grant for girls' education activities; both grants were Tanzania shillings (TZS) 550,000.⁵
- Distribution of school noticeboards and support materials.
- Training of community facilitators by civil society organisations (CSOs) to support the development of Community Education Needs Assessments (CENAs).
- Training on business plan development for head teachers / teachers and community business leaders; provision of income-generation activity (IGA) grants of TZS 1,500,000⁶ to 50% of schools.

More schools are holding annual meetings with parents and teachers than at baseline, a key communication and engagement mechanism. There has been an improvement in the number of SCs that have been trained on their roles and responsibilities. However, a sizeable minority of SCs are still not meeting regularly and head teachers do not feel the support from SCs is good.

The PTP initiative has had some success: almost all schools have a PTP, and most have had training. However, PTPs are not very active in terms of meetings, and a sizeable minority did not take any action to improve education in 2017. The activities that did take place were most commonly to improve pupil attendance and punctuality, although pupil attendance did not change between midline and endline and actually worsened slightly for boys. PTPs do not yet appear to be bridging the gap between schools and the community: most parents who are not on the PTP do not know about them. Similarly, noticeboards were distributed to most but not all schools, yet only half of schools are displaying them in a public place, showing no improvement since baseline. Only a minority of parents are aware of the noticeboards and use them for information. Thus, these initiatives have not substantially improved communication between schools and parents, and this raises the question of whether they are relevant to the context, or whether more effort is needed to make them effective.

EQUIP-T's grants to schools have been effectively administered: the targeted numbers of schools received the first PTP grant and IGA grant, largely in the correct amount, and most have spent the first PTP grant by endline. Most IGA recipient schools and communities have started implementing the IGAs, which is a promising sign for sustainability, as, if they are successful, in future the additional funds could be used to continue school improvement activities started under EQUIP-T.

⁵ Equivalent to £187 or US\$240 at current exchange rates.

⁶ Equivalent to £510 or US\$655 at current exchange rates.

Conducive learning environments for marginalised children

Summary of implementation under Component 4B: Conducive learning environments for marginalised children, particularly girls and children with disabilities

- Training to enable schools to set up *Jiamini Uwezo Unao* (JUJ) ('Be confident, you can') clubs for pupils in upper standards. JUJ clubs are non-academic and non-extracurricular school clubs that 'provide a platform for children, especially girls, to receive guidance on managing the barriers to education, to discuss their opinions and to act as change-makers in their school and communities.' (EQUIP-T Managing Agent (MA), 2017).
- Provision of a second PTP grant for activities to support girls' education (as under component 4A).
- Distribution of the *Shujaaz* magazine to schools and communities as part of a behaviour change campaign related to girls' education.

The impact evaluation confirms the importance of the programme's efforts to support marginalised or vulnerable children. A large minority of pupils come from households that are below the poverty line, most children did not eat before coming to school, almost a third have at least a 45-minute walk to get to school, and most children do not speak Kiswahili at home. Across various measures, pupils in the EQUIP-T districts are characterised by many dimensions of vulnerability and it is important that the programme has recognised and responded in ways to overcome these barriers.

There has been relative success in rolling out the new conducive learning environments initiatives. Most schools have been trained on and now have JUJ clubs, which are relatively active in meeting, and the majority carried out an activity in 2017, though most did not carry out activities to target specific groups of vulnerable pupils. Most schools have received the PTP girls' education grant, but only a quarter have spent it so far. Most schools have received the *Shujaaz* magazine. Whilst it is harder to measure whether these outputs have led to the anticipated extra support for girls and other marginalised children, the survey finds that schools have not yet integrated these issues into activities in their SDPs.

Recommendations

Recognising that the programme will close in January 2020, this impact evaluation makes recommendations for the EQUIP-T MA that focus largely on reinforcing the existing interventions, rather than introducing new initiatives. Recommendations for the Government of Tanzania in deciding how to take forward the learning from the EQUIP-T programme are interim at this stage. After the qualitative and costing study parts of the endline, recommendations for the Government on scalability will be finalised.

For the EQUIP-T MA, the impact evaluation recommends running refresher or mop-up training in 2019 to ensure that a much higher share of the intended beneficiaries have received all the training. This applies to in-service training for early grade teachers, SLM training, and training for WEOs, SCs, and PTPs. With the focus on training, the programme, together with its local government partners, should consider ways to overcome some of the challenges that are reported to be contributing to irregular attendance, such as the timing and duration of sessions, the location of training venues, and ensuring full coverage of out-of-pocket expenses. Refresher teacher training should emphasise the use of TLMs in lessons. EQUIP-T should also consider developing or adapting modules to support teachers to better deal with varied class sizes and with language barriers for children who do not speak Kiswahili at home. The programme should work with PO-RALG to review and consider simplifying the templates used for SIS data entry, to make the system more accessible and ultimately more useful for school management. Finally, EQUIP-T should find ways to support its local government partners to strengthen the relationship-building role of the PTP for the wider community.

For the Government, the impact evaluation recommends a continued focus on mitigating external contextual challenges, particularly the high turnover among education professionals, large class sizes (recognising that significant construction efforts designed to alleviate classroom shortages are already underway), and providing support to teachers to deal with language barriers for children who do not speak Kiswahili at home. The Government could also look to address the high levels of teacher school and classroom absenteeism through the teacher management element of the Agency for the Development of Education Management's (ADEM's) SLM materials and training. The same training could also usefully focus on equipping head teachers with strategies to reduce pupil absenteeism, which has remained very high over the evaluation period.

If the Government of Tanzania adopts elements of EQUIP-T's in-service training, it should also consider ways to encourage greater attendance of teachers, and in addition to training for SCs, look for ways to encourage SCs to be more active in their meetings and in the support they provide to schools. Finally, the evaluation evidence so far shows that PTPs and noticeboards have been relatively successfully rolled out but are not widely known about by parents. Thus, the Government of Tanzania should be cautious in considering scaling up these initiatives with their current design if the main objective is to improve communication between schools and the wider parental body and community. PTPs may, however, serve other functions in terms of support in the classroom.

Table of contents

Acknowledgements	i
Executive summary	ii
List of figures, tables, and boxes	xiv
List of abbreviations	xvi
Part A: Impact evaluation objectives, background, and methods	1
1 Impact evaluation objectives and EQUIP-T overview	2
1.1 Purpose of this report	2
1.2 EQUIP-T design and theory of change, and scope of the impact evaluation	3
1.2.1 EQUIP-T design and overarching theory of change	3
1.2.2 Scope of the impact evaluation and theory of change	4
1.3 EQUIP-T implementation progress	5
1.4 Education sector policy context	6
1.4.1 Education policy changes since baseline	6
1.4.2 Other large education programmes since baseline	7
2 Impact evaluation objectives, methods and endline design	8
2.1 Impact evaluation objectives	8
2.2 Impact evaluation design and methods	8
2.3 Endline design	9
2.3.1 Sequencing of mixed methods	9
2.3.2 Research priorities for the quantitative endline evaluation	10
2.4 Endline survey instruments and actual samples	11
2.5 How to read the report	12
2.5.1 Impact estimates	12
2.5.2 Descriptive trends in programme areas	14
2.5.3 Red, amber, green (RAG) summary results tables	14
2.5.4 Terminology	15
2.6 Structure of Volume I	15
Part B: Findings	17
3 Pupil learning, background, and home support	18
3.1 Overall programme goals and expectations of change in pupil learning by endline	18
3.2 Trends in pupil backgrounds and home support for education in treatment areas	20
3.2.1 Has the personal and economic profile of Standard 3 pupils changed since baseline?	20
3.2.2 Has the home support for education and demand for schooling for Standard 3 pupils changed since baseline?	21
3.2.3 Summary of trends in the profile of Standard 3 pupils	23
3.3 Trends in early grade learning and impact of EQUIP-T	23
3.3.1 How has early grade pupil learning changed overall in treatment areas?	24
3.3.2 What is the impact of EQUIP-T on early grade pupil learning?	27
3.3.3 Why have there been gains in pupil learning overall, and what are the most likely explanations for EQUIP-T's impact?	30
3.4 Trends in early grade learning gaps by gender in programme areas	32
3.5 Trends in early grade learning gaps by home language in programme areas	33

3.6	Trends in learning gaps by household poverty status in programme areas	34
3.7	Summary of findings on pupil learning	36
3.7.1	Impact results	36
3.7.2	Trends in EQUIP-T programme areas	36
4	Teacher performance	38
4.1	Teacher performance: theory of change, implementation, and expectations of change	38
4.1.1	Theory of change	38
4.1.2	Implementation since baseline	39
4.1.3	Expected changes	41
4.2	Findings: EQUIP-T in-service training for teachers and COL	42
4.2.1	Have teachers received EQUIP-T in-service training? (EQUIP-T input)	42
4.2.2	Teacher attendance at EQUIP-T in-service training (EQUIP-T input to output assumption)	45
4.2.3	Relevance and accessibility of EQUIP-T in-service training (EQUIP-T input to output assumption)	45
4.2.4	Are teacher COL structures operating? (EQUIP-T output)	46
4.3	Findings: teacher capacity and confidence	46
4.3.1	Has teacher capacity and confidence improved? (EQUIP-T output)	46
4.4	Findings: class size, teacher turnover, and job satisfaction (EQUIP-T output to intermediate outcome assumptions)	47
4.4.1	Class size	47
4.4.2	Level of teacher turnover	48
4.4.3	Teacher job satisfaction and motivation	49
4.5	Findings: classroom teaching	50
4.5.1	Has the use of inclusive and gender-responsive teaching practices in the classroom increased? (EQUIP-T intermediate outcome)	51
4.5.2	Have teaching and assessment practices improved? (EQUIP-T intermediate outcome)	52
4.6	Findings: provision, availability, and use of EQUIP-T TLMs	56
4.6.1	Have schools received EQUIP-T TLMs? (EQUIP-T input)	56
4.6.2	Has the availability of TLMs in classrooms increased? (EQUIP-T output)	56
4.6.3	Has the use of TLMs in classrooms increased? (EQUIP-T intermediate outcome)	57
4.7	Findings: instructional time and teacher attendance	57
4.7.1	Teacher absence from school and classrooms, and punctuality	57
4.7.2	Has instructional time increased? (EQUIP-T intermediate outcome)	60
4.8	Summary of findings on teacher performance	61
5	SLM	64
5.1	SLM: theory of change, implementation, and expectations of change	64
5.1.1	Theory of change	64
5.1.2	Implementation since baseline	65
5.1.3	Expected changes	65
5.2	Findings: head teacher background and turnover	66
5.2.1	Head teacher background characteristics and turnover	66
5.3	Findings: EQUIP-T training for head teachers	67
5.3.1	Has EQUIP-T provided SLM in-service training for head teachers? (EQUIP-T input)	67
5.3.2	Has EQUIP-T provided early grade teacher in-service training to head teachers? (EQUIP-T input)	70
5.4	Findings: head teacher capacity and SLM practices	70

5.4.1	Has head teacher capacity changed? (EQUIP-T output)	70
5.4.2	Have head teachers' SLM practices changed? (EQUIP-T intermediate outcome)	71
5.5	Findings: head teacher absence and job satisfaction	74
5.5.1	Head teacher absence and job satisfaction (Equip-t output to intermediate outcome assumptions)	74
5.6	Findings: capitation grant payments and in-kind resources	75
5.6.1	Capitation grant payments received in full and in-kind resources received (EQUIP-T output to intermediate outcome assumption)	75
5.7	Findings: SIS training and use	76
5.7.1	Has EQUIP-T trained head teachers on SIS and provided SIS tablets? (EQUIP-T input)	76
5.7.2	Is the EQUIP-T SIS functional? (EQUIP-T output)	77
5.7.3	EQUIP-T SIS tablet software is fit for purpose and tablet-based SIS is appropriate (EQUIP-T input to output assumption)	78
5.7.4	Does the SIS provide useful reports to support SLM? (EQUIP-T intermediate outcome)	78
5.8	Findings: WEO support for schools and head teachers	79
5.8.1	Is WEO support for head teachers effective? (EQUIP-T output, component 3B)	79
5.8.2	Have WEO school support practices changed? (EQUIP-T intermediate outcome)	79
5.8.3	WEO turnover (EQUIP-T input to output assumption)	80
5.8.4	Are head teachers reporting to WEOs / districts? (EQUIP-T output)	80
5.8.5	Are head teachers attending ward education and COL meetings? (EQUIP-T output)	81
5.9	Summary of findings on SLM	81
6	Community participation and accountability	84
6.1	EQUIP-T component 4A theory of change, implementation, and expectations of change	84
6.1.1	Theory of change	84
6.1.2	Implementation since baseline	85
6.1.3	Expected changes	85
6.2	Findings: SCs	86
6.2.1	Has EQUIP-T provided training for SCs? (EQUIP-T input)	86
6.2.2	Has SC capacity increased? (EQUIP-T output)	86
6.3	Findings: PTPs	87
6.3.1	Has EQUIP-T provided training for PTPs? (EQUIP-T input)	87
6.3.2	Has PTP capacity increased and are PTPs active? (EQUIP-T output)	88
6.3.3	Have PTPs taken action to improve education, and are parents' engaged? (EQUIP-T intermediate outcome)	89
6.4	Findings: PTP and IGA grants	89
6.4.1	Has EQUIP-T provided PTP grants (EQUIP-T input) and have PTP grants been spent? (EQUIP-T output)	89
6.4.2	Has EQUIP-T provided IGA training and grants (EQUIP-T input) and have IGAs started? (EQUIP-T output)	90
6.5	Findings: school notice boards	91
6.5.1	Has EQUIP-T provided school notice boards? (EQUIP-T input)	91
6.5.2	Are school notice boards publicly accessible and used? (EQUIP-T output)	91
6.5.3	Are parents aware of and reading school notice boards? (EQUIP-T intermediate outcome)	92
6.6	Findings: CENAs	92
6.6.1	Was a CENA undertaken and actions taken based on it? (EQUIP-T input and intermediate outcome)	93

6.7	Has schools' interaction with all parents and the wider community improved? (EQUIP-T intermediate outcome)	93
6.8	Summary of community 4A evidence	94
7	Conducive learning environments for marginalised children, particularly for girls and children with disabilities	97
7.1	EQUIP-T component 4B: theory of change, implementation, and expectations of change	98
7.1.1	Theory of change	98
7.1.2	Implementation since midline	99
7.1.3	Expected changes	99
7.2	Profile of vulnerability in programme schools	99
7.3	Findings	101
7.3.1	Have schools received the various inputs under this component as intended? (EQUIP-T input)	101
7.3.2	Are JUU clubs established, active, and supporting marginalised children? (EQUIP-T output and intermediate outcome)	101
7.3.3	Has the PTP girls' education grant been spent, and on what? (EQUIP-T output and intermediate outcome)	102
7.3.4	Have inputs related to child protection and positive behaviour management been received and are schools promoting child protection and anti-violence? (EQUIP-T input, output, and intermediate outcome)	103
7.3.5	Are teachers using inclusive teaching practices in the classroom to support marginalised children? (EQUIP-T intermediate outcome)	103
7.3.6	Are there inclusive strategies at the school level that support marginalised children? (EQUIP-T intermediate outcome)	105
7.4	Summary of findings on conducive learning environments	106
PART C: Conclusions, recommendations, and lessons		109
8	Conclusions, recommendations, and lessons	110
8.1	Conclusions on impact	110
8.2	Conclusions on effectiveness, relevance, and sustainability	110
8.2.1	Targeting improvements in pupil learning	111
8.2.2	Teachers' performance	112
8.2.3	SLM	113
8.2.4	Community participation	114
8.2.5	Conducive learning environments for marginalised children	114
8.3	Recommendations	115
8.3.1	Recommendations for EQUIP-T MA	115
8.3.2	Recommendations for Government	116
8.4	Lessons learned	117
8.5	Implications for final phase of the impact evaluation	118
References		119
Annex A	Agreed terms of reference	124
A.1	Impact evaluation purpose in original TOR	124
A.2	Impact evaluation design options	124
A.3	DFID design choices	125
A.4	What the impact evaluation will measure under the agreed terms of reference	125
A.5	Evaluation questions	127
A.6	Revised purpose of the impact evaluation	128
A.7	Changes to the impact evaluation design since the technical proposal	128
Annex B	Supplementary information on EQUIP-T	130

B.1	Constraints acting on children’s ability to learn	130
B.2	Updated EQUIP-T theory of change	130
B.2.1	Evolution of EQUIP-T’s theory of change	131
B.2.2	Assumptions and explanations behind the new elements in the theory of change	132
B.3	EQUIP-T implementation progress between baseline and end-March 2018	133
B.4	EQUIP-T in-service training implementation details	137
B.5	EQUIP-T SLM implementation details	138
B.6	EQUIP-T logframe indicators	139
Annex C	Endline evaluation matrix, part I	141
Annex D	Education policy context and other programmes	145
	Education policy changes since baseline	145
Annex E	Teaching practices descriptors	147

List of figures, tables, and boxes

Figure 1: EQUIP-T programme districts (prior to the extension in 2017) and the impact evaluation district sample ¹	9
Figure 2: Classification of indicator values in summary results tables	14
Figure 3: High-level results chain from theory of change ^{1, 2}	18
Figure 4: Trends in pupils' economic background (trends in programme areas)	21
Figure 5: Distribution of pupil learning in Kiswahili by performance band in treatment areas, baseline, midline, and endline (%) (trends in programme areas)	25
Figure 6: Distribution of pupil learning in maths by performance band in treatment areas, baseline, midline, endline (%) (trends in programme areas)	26
Figure 7: Main results chain for EQUIP-T component 1A teacher performance	39
Figure 8: EQUIP-T's current model of in-service training for teachers	39
Figure 9: EQUIP-T school-based in-service training implementation since baseline (trends in programme areas)	44
Figure 10: Class sizes by standard: baseline, midline, and endline (trends in programme areas)	48
Figure 11: Observed teaching practices at the start and end of a lesson (trends in programme areas)	53
Figure 12: Observed teaching practices during the middle stages of a lesson (trends in programme areas)	54
Figure 13: Actual instructional time for Kiswahili and mathematics for Standards 1 and 2 (trends in programme areas)	60
Figure 14: Main results chain for EQUIP-T component 2, SLM	65
Figure 15: Contents of EQUIP-T SLM in-service training ever received	69
Figure 16: Reasons for schools not entering data on pupil enrolment and current teachers in the SIS77	78
Figure 17: Difficulties in using SIS tablets	78
Figure 18: Main results chain for EQUIP-T component 4A community	85
Figure 19: Main topic discussed during SC meetings (trends in programme areas)	87
Figure 20: Accessibility and use of school notice boards at endline	92
Figure 21: Main results chain for EQUIP-T component 4B conducive learning environments	98
Figure 22: Types of activities carried out by JUU clubs in 2017	102
Figure 23 Constraints on children's capability to learn to their full potential	130
Figure 24 Updated EQUIP-T theory of change (EQUIP-T MA)	131
Table 1: EQUIP-T programme components and sub-components from mid-2017	3
Table 2: Endline quantitative survey sample sizes	11
Table 3: Trends in demand for schooling, supply response, and pupil absenteeism (trends in programme areas)	23
Table 4: Proportion of pupils in bottom and top performance bands for Kiswahili and maths at baseline, midline, and endline by gender (%) (trends in programme areas)	32
Table 5: Proportion of pupils in bottom and top performance bands for Kiswahili and maths at baseline, midline, and endline by home language (%) (trends in programme areas)	34
Table 6: Proportion of pupils in bottom and top performance bands for Kiswahili and maths at baseline, midline, and endline by poverty status (%) (trends in programme areas)	35
Table 7: Gender balance in teachers' interactions with pupils in the classroom (trends in programme areas)	51
Table 8: Spatial inclusiveness of teachers' interactions with pupils in the classroom (trends in programme areas)	51
Table 9: Teacher absenteeism and punctuality on the day of the survey (trends in programme areas)	58
Table 10: Summary of evidence on teacher performance	61
Table 11: EQUIP-T SLM in-service training received, by head teachers' turnover status (trends in programme areas)	68
Table 12: Availability of SDPs (trends in programme areas)	71

Table 13: Teacher management practices reported by Standards 1 to 3 teachers (trends in programme areas)	72
Table 14: Selected school characteristics and infrastructure (trends in programme areas)	76
Table 15: Summary of evidence on SLM and SIS	81
Table 16: PTP grant 1 spending	90
Table 17: Summary of evidence on community participation and accountability	94
Table 18: Pupil absenteeism from school by certain vulnerability dimensions	100
Table 19: Summary of evidence on conducive learning environments for marginalised children	106
Table 20 What the impact evaluation will measure	127
Table 21 Implementation of sub-component 1A teacher performance	134
Table 22 Implementation of sub-components 2A and 2B school leadership and management	135
Table 23 Implementation of sub-component 4A community participation and accountability	136
Table 24 Implementation of sub-component 4B conducive learning environments for girls, children with disabilities and other marginalised children	137
Table 25: Endline evaluation matrix part I (partial, related to quantitative evidence)	141
Table 26 Teaching practices and descriptors	147
Box 1: Large-scale primary education development programmes, apart from EQUIP-T	7
Box 2: Assumptions related to the interpretation of the impact estimates	13
Box 3 Key terminology used in this report	15
Box 4: Learning achievement targets from programme logframes, and original intervention summary	19
Box 5: EQUIP-T's SRP	22
Box 6: EQUIP-T impact on early grade pupil learning	27
Box 7: Description of key concepts for EQUIP-T teacher component	38
Box 8: Staged implementation of in-service training and delivery of TLMs	41
Box 9: EQUIP-T's impact on teachers' absence from school and classrooms	59
Box 10: Description of key EQUIP-T SLM concepts	64
Box 11: Level of head teacher turnover ¹	66
Box 12: Key school characteristics and infrastructure	75
Box 13: Description of key community concepts and partners	84
Box 14: Has EQUIP-T provided training on SCs, PTPs, grants, and business plans for head teachers? (EQUIP-T input)	88
Box 15: Key concepts for EQUIP-T conducive learning environments component	97
Box 16: Component 4B activities planned from mid-2018 onwards	99
Box 17 List of EQUIP-T logframe indicators related to the components covered by the first part of the endline evaluation	139

List of abbreviations

3Rs	Reading, writing, and arithmetic
ADEM	Agency for the Development of Education Management
ATT	Average treatment effect on the treated
BL	Baseline
BRN-Ed	Big Results Now in Education
CAPI	Computer aided personal interviewing
CENA	Community Education Needs Assessment
CSO	Civil society organisation
DFID	Department for International Development
DID	Difference-in-differences
DSI	District school inspector
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EPforR	Education Program for Results
EQUIP-T	Education Quality Improvement Programme in Tanzania
ESDP	Education Sector Development Plan
ETP	Education and Training Policy
IDELA	International development and early grade learning assessment
IE	Impact evaluation
IGA	Income-generating activity
INCO	In-service training coordinator
ISS	Institutional strengthening and sustainability
JUU	<i>Jiamini Uwezo Unao</i> ('Be confident, you can')
LANES	Literacy and Numeracy Education Support Programme
LGA	Local Government Authority
MA	Managing Agent
ML	Midline
MOEST	Ministry of Education, Science and Technology

MOEVT	Ministry of Education and Vocational Training ⁷
OECD-DAC	Organisation for Economic Co-operation and Development – Development Assistance Committee
OPM	Oxford Policy Management
PO-RALG	President’s Office Regional Administration and Local Government
PSM	Propensity score matching
PTP	Parents–teachers partnership
RCT	Randomised control trial
RTI	Research Triangle International
SC	School Committee
SDP	School Development Plan
SIDA	Swedish Development Agency
SIS	School information system
SLM	School leadership and management
SPMM	School performance management meeting
SRP	School Readiness Programme
SQA	School quality assurance/assurer
STEP	Student Teacher Enrichment Program
TCF	Teacher Competency Framework
TDNA	Teacher development needs assessment
TLMs	Teaching and learning materials
TOR	Terms of reference
TZS	Tanzanian shilling
UNESCO	UN Educational, Scientific and Cultural Organization
UNICEF	UN Children’s Fund
URT	United Republic of Tanzania
USAID	US Agency for International Development
WEO	Ward Education Officer

⁷ Before the change of government in 2015 MOEST was called the Ministry of Education and Vocational Training (MOEVT).

Part A: Impact evaluation objectives, background, and methods



1 Impact evaluation objectives and EQUIP-T overview

1.1 Purpose of this report

This report presents the findings from the first part of the endline round of the impact evaluation of the Education Quality Improvement Programme in Tanzania (EQUIP-T). This follows two previous rounds of research—a baseline in 2014 and a midline in 2016. This is an independent evaluation which is overseen by a Reference Group that is chaired by the Commissioner for Education.⁸ The independent evaluation is being carried out by Oxford Policy Management (OPM).

EQUIP-T began in 2014 as a four-year, Government of Tanzania programme funded by the UK Department for International Development (DFID). Under this four-year programme, activities were initially targeted at five, and later seven, of the most educationally disadvantaged regions in Tanzania. The aim of the programme, which has not changed since the outset, is to increase the quality of primary education and improve pupil learning outcomes, in particular for girls.⁹ The programme planned to reach about 2.3 million primary pupils, about one-quarter of total primary enrolment in Tanzania, under a budget of approximately £50 million.

In 2017, the programme was extended by a year and a half to January 2020 and the budget increased to approximately £90m. As well as continuing activities in the original seven regions, some new sub-components were added in these regions, and a reduced package of interventions was planned for two new regions. The extension of EQUIP-T had implications for the endline design of the impact evaluation, which was originally anticipated to be completed in full in 2018. Some trade-offs had to be made in scope and methods (explained further in Chapter 2, Section 2.3), with the result that this first part of the endline evaluation is focused on estimating the casual impact of the programme using quantitative methods. Qualitative endline research, which will serve to deepen understanding of key quantitative findings, and a costing study, will follow in 2019. The full endline evaluation report, due in early 2020, will synthesise findings from all of these evaluation components.

The main aims of this first component of the endline evaluation are to estimate the impact of EQUIP-T on pupil learning achievement, and to assess the effectiveness of the school- and community-level EQUIP-T interventions, after nearly four years of implementation (44 months).¹⁰ The results are intended to inform further adjustments to the programme before it finishes in January 2020, as well as to promote accountability and lesson learning for DFID and the Government of Tanzania. Its findings will also help to guide the design of the qualitative endline research in 2019.

The main data source for this first component of the endline evaluation is three rounds of a quantitative survey from a panel of 200 primary schools (100 EQUIP-T treatment schools, and 100 control schools) that started with a baseline in 2014.

The scope of this quantitative part of the endline evaluation was discussed with DFID and members of the Reference Group in the first quarter of 2018. The results of the discussion are set out in the approved Endline Planning Report: Part I (OPM, 2018); this provides the guidance

⁸ The Reference Group's membership comprises government officials from different ministries, departments, and agencies with responsibility for education, academics in the field of education, and members from education research organisations, the EQUIP-T Managing Agent (MA), other large education development programmes, and DFID.

⁹ There are 26 regions in mainland Tanzania.

¹⁰ To the extent that this is possible using quantitative methods. Some of the anticipated changes related to the programme components require qualitative research methods (for example, improvement in accountability).

document for the analysis presented in this report. The agreed terms of reference (TOR) for the full impact evaluation are given in Annex A, although these have not been adjusted to take account of the extension to the programme.

This Endline Evaluation Report (Part I) is organised into two volumes:

- Volume I, Results and Discussion: presents the main findings and discusses the key recommendations for consideration by the programme's MA, and the Government of Tanzania, as well as broader lessons.
- Volume II, Methods and Supplementary Evidence: contains the detailed methods sections on the survey design and sampling, impact estimation, and measurement of pupil learning. It also describes some sets of supplementary results to support the conclusions reached in Volume I. This volume provides important statistical reference tables, which contain all the estimates presented in Volume I, as well as a large number of supplementary indicators.

1.2 EQUIP-T design and theory of change, and scope of the impact evaluation

1.2.1 EQUIP-T design and overarching theory of change

Following its extension in 2017, EQUIP-T comprises 10 sub-components, set out in Table 1.

Table 1: EQUIP-T programme components and sub-components from mid-2017

Component and sub-components ¹	Scope of impact evaluation	Scope of quantitative endline evaluation
1 Improved access to quality education		
1A Teacher performance	✓	✓
1B School Readiness Programme (SRP) and satellite schools (non-construction)	✓ partly ²	X
1C Classroom and satellite school construction	X	X
2 Strengthened school leadership and management (SLM)		
2A SLM capacity building	✓ partly	✓ partly
2B School information system (SIS)	✓ partly	✓ partly
3. Strengthened district planning and management		
3A District capacity building	✓	X ³
3B Local Government Authority (LGA) (district) grant monitoring	✓	X
4 Stronger community participation and demand for accountability in education		
4A Community participation and accountability	✓	✓ partly
4B Conducive learning environment for marginalised children, particularly for girls and children with disabilities	✓ partly	✓ partly
5 Improved learning and dissemination		
5 Learning and dissemination	X	X

Source: OPM 2018, p. 13. Notes: (1) The five original components are in bold. The rest are new components under the extension. (2) This only relates to SRP, and its coverage is very limited. (3) There is one exception: the school visits and support that Ward Education Officers (WEOs) provide to schools are included in the scope of the quantitative endline evaluation.

The five original sub-components are highlighted in bold in the table. These were originally designed to overcome a set of key constraints that EQUIP-T identified as undermining pupils'

capability to learn to their full potential, in disadvantaged parts of Tanzania (Annex B.1). The objectives of the original components are to: improve the performance of teachers (1A); strengthen SLM (2A); enhance district planning and management (3A); strengthen community participation and demand for accountability (4A); and improve learning and dissemination of results (5). Overall, the emphasis of this first set of EQUIP-T interventions is on strengthening the education system to deliver *quality* education.

At the time of the extension design, EQUIP-T identified that access to education, particularly for remote and marginalised children, is a more serious problem than had initially been understood. This includes barriers to enrolling in pre-school and primary school, and to participating and being successful once enrolled. With the aim of mitigating the effects of these barriers, EQUIP-T added five new components (Table 1), including 4B, which aims to ensure an inclusive learning environment. Some of the initiatives under this sub-component build on earlier activities under the original components that were designed to promote gender and social inclusion.

The programme's overarching theory of change conceptualises the components as mutually reinforcing in overcoming the identified barriers at local, school, and district/national levels. Taken together, these components are expected to lead to better quality education, especially for girls (EQUIP-T outcome), and to improved learning outcomes, especially for girls across Tanzania (EQUIP-T impact). To support national adoption and scale-up of successful parts of the programme, EQUIP-T has an institutional strengthening and sustainability (ISS) strategy integrated into its theory of change. The EQUIP-T MA's updated theory of change is provided in Annex B.2, together with a summary of the new sub-components (1B, 1C, 2B, and 4B).

1.2.2 Scope of the impact evaluation and theory of change

The original scope of the impact evaluation includes four of the original five components, namely: teacher performance (1A), SLM (2A), district planning and management (3A), and community participation and accountability (4A). Fully absorbing the new sub-components into the evaluation at this endline stage was not possible, but an expanded scope was agreed in the Endline Planning Report Part I (OPM, 2018). Table 1 shows that it is only the construction programme (1C) and the learning and dissemination component (5) that are completely excluded from the scope. The other new sub-components are partly included, which means that the evaluation will produce some evidence (quantitative or qualitative) related to these sub-components.

This first part of the endline evaluation is based on quantitative data from the panel survey of schools. This narrows the scope of this part of the evaluation to the sub-components marked in the final column of Table 1, because some of the expected changes from the components are not able to be measured well using quantitative methods.¹¹ The qualitative component of the evaluation will follow in 2019, and its themes will partly depend on the results from this study.

The design of this first part of the endline evaluation is based on the detailed theory of change for the sub-components in scope. By working with the EQUIP-T programme staff in early 2018, it was possible to document the detailed results chains¹² implicit in the overarching theory of change, including the key assumptions underpinning the links.¹³ This was turned into the Endline

¹¹ For instance, the community component 4A is aiming to strengthen accountability relationships between schools and communities—this requires qualitative research to understand change well.

¹² A results chain is the path from the receipt of inputs to delivery of outputs, and contributions to intermediate outcomes and impact.

¹³ This was not exhaustive, covering all results-chains embedded in the theory of change, but instead focused on the main interventions within the scope of the evaluation, expected changes to follow from these, and key assumptions.

Evaluation Matrix Part I (OPM, 2018, and Annex C) containing the endline evaluation questions, linked to the theory of change, that can be answered using quantitative evidence. The present report presents findings on expected changes and assumptions at each level of the result chains (see Chapters 4 to 7), as evidence of reasons for programme impact on pupil learning outcomes (discussed in Chapter 3).

In terms of geographical scope, the impact evaluation covers the first five regions in which EQUIP-T started field implementation in August 2014: Dodoma, Kigoma, Shinyanga, Simiyu, and Tabora. Comments on the external validity (expected scope of generalisability) of the evaluation findings are given in Section 2.4 in Chapter 2.

1.3 EQUIP-T implementation progress

To give a flavour of what EQUIP-T has been supporting in schools and communities for the sub-components that are in focus for this part of the endline evaluation, this section briefly summarises the main interventions since baseline. Tables containing more details of implementation progress until the endline survey in April 2018 are given in Annex B.3. Chapters 4 to 7, which cover findings on each of the sub-components in turn, give further description of the activities.

- **Improving the performance of teachers (1A):** Delivery of in-service training for early grade teachers on: Kiswahili (literacy), the new Standards 1 and 2 curriculum (see Section 1.4.1 for details), maths (numeracy), and gender-responsive pedagogy; and distribution of teaching and learning materials (TLMs) for early grade pupils and teachers.
- **Enhancing SLM (2A and 2B):** Delivery of in-service training for head teachers and WEOs on SLM roles and responsibilities, school development planning, and school performance management meetings (SPMMs); provision of tablets to head teachers accompanied by initial training on the SIS.
- **Strengthening community participation and accountability (4A):** Supply of noticeboards to schools; support for communities to develop education needs assessments; training of school committees (SCs) on roles and responsibilities; support for set up, and training of, parent–teacher partnership bodies (PTPs); disbursement of two PTP grants; training of selected teachers and community members on business plan development; and disbursement of income-generating activity (IGA) grants to 50% of schools in each district.
- **Ensuring a conducive learning environment for girls and marginalised children (4B):** Support for setting up *Jiamini Uwezo Unao* (JUJ) ('Be confident, you can') clubs for pupils in upper standards to promote pupil welfare and inclusion; distributing a magazine called *Shujaaz* to schools as part of a behaviour change campaign related to girls' education; and, linked to the PTP grants under 4A, the second grant is intended for activities to support girls' education.

It is important to highlight that EQUIP-T introduced a decentralised funding mechanism for its programme support funds (approximate budget of £36 million¹⁴) in the 2015/16 financial year. Almost all the activities described above, bar centrally procured materials and contracts and region-level training, are funded by districts using programme funds managed through the Government of Tanzania's public financial management system. This is one of the reasons that implementation timing, modality, and intensity can vary across districts, within the guidelines provided by EQUIP-T. Much of the programme comprises capacity building via training

¹⁴ Pre-extension, the total budget for EQUIP-T was £50 million, of which £36 million was for programme support activities.

programmes, and while EQUIP-T provides guidelines on the target participants, the actual participants are decided by schools, WEOs, and/or districts.

The decentralised funding mechanism is a key part of EQUIP-T's ISS strategy, giving hands-on experience to district officials in planning, funding and managing EQUIP-T activities, under the strategic guidance of regional officials (supported by EQUIP-T regional teams). Over time, EQUIP-T has purposefully reduced its direct involvement at school level, increasingly leaving the co-ordinating and leadership role to Government officials at different levels.

Implementation of some EQUIP-T components has been slower than anticipated (for example, see notes under the tables in Annex B.3). One of the contributing factors, according to the programme's MA, has been delays in Government approval of materials and fund release for districts.¹⁵

1.4 Education sector policy context

1.4.1 Education policy changes since baseline

Since the baseline research in 2014 there have been at least four major national policy changes that affect primary education across the country (see below, and Annex D for more details). Recognising the changing education context since the baseline is very important for the impact evaluation because national trends rather than EQUIP-T could be driving some of the observed changes. The impact evaluation methodology (outlined in Chapter 2) ensures that national trends do not confound the estimates of programme impact, but they need to be considered when interpreting descriptive trends found in EQUIP-T treatment schools.

- **Fee-free basic education policy:** From December 2015, the Government of Tanzania announced that no fees or other compulsory contributions would be charged for pupils from Standard 1 to Form 4.¹⁶ The directive did not explicitly mention pre-primary, but the latest Education Sector Development Plan (ESDP) (Ministry of Education, Science and Technology (MOEST), 2018) makes it clear that one year of pre-primary education is also fee-free.¹⁷
- **Monthly school capitation grants:** From January 2016, the Government of Tanzania started paying monthly capitation grants directly into schools' bank accounts, rather than via the district authorities as under the previous system. This payment is intended to compensate schools for their loss of parental contributions.
- **New competency-based early grade curriculum:** From part-way through the 2015 school year, the Government introduced a new competency-based curriculum for Standards 1 and 2 that focuses on reading, writing, and arithmetic (3Rs), rather than a larger set of subjects. It promotes a new phonics-based approach to teaching children to read. In 2016, the Government introduced a new curriculum for Standards 3 and 4, which follows on from the new Standards 1 and 2 curriculum in approach but has more advanced content.
- **Human resource management:** From 2017, the Government introduced higher minimum qualifications for head teachers (diplomas) and WEOs (degrees), as part of a professionalisation policy. This means that under-qualified head teachers and WEOs had to

¹⁵ This information comes from the EQUIP-T MA's written feedback comments on the first draft of this report (received by the evaluation team 15 November 2018).

¹⁶ Government Circular Number 5 of 2015 reads: 'Provision of free education means pupils or students will not pay any fee or other contributions that were being provided by parents or guardians before the release of the new circular'.

¹⁷ The executive summary states that the Government of Tanzania is committed to '12 years of free and compulsory Basic Education to the entire population'. This includes one year of pre-primary.

leave their positions. In the same year, the Government also dismissed a large number of civil servants, including teachers, for having fake qualifications. More generally, from early 2016, the new Government set a very high-profile national agenda to encourage public servants (including education professionals) to work hard, and to carry out their duties professionally, to improve public services.

1.4.2 Other large education programmes since baseline

Alongside the implementation of EQUIP-T, which started in 2014, there have been several large-scale primary education development programmes working to improve the quality of primary education, under the umbrella of the Government's ESDP. Although interventions differ across these programmes, they share at least one common objective which is to improve early grade pupil learning. The three main programmes are summarised in Box 1.

Box 1: Large-scale primary education development programmes, apart from EQUIP-T

Literacy and numeracy education support programme (LANES): This runs from 2014/15 to 2018/19, funded by the Global Partnership for Education, with a budget of US\$95 million. It has national coverage (26 mainland regions), except for a few activities.

Education Program for Results (EPforR), formerly Big Results Now-Education (BRN-Ed): This runs from 2014/15 to 2020/21, funded by World Bank, Swedish International Development Cooperation Agency (SIDA), DFID, and the Government of Tanzania, with a budget of US\$416 million. It has national coverage, except for a few of the earlier activities under **BRN-Ed**.

Tusome Pamoja ('Let's read together'): This runs from 2016 to 2021, funded by the US Agency for International Development (USAID). It covers four mainland regions (Iringa, Morogoro, Mtwara, and Ruvuma) and all districts in Zanzibar.

Source: OPM 2018

From the perspective of the impact evaluation of EQUIP-T, it is important to understand whether any of these programme's activities have taken place in the EQUIP-T districts or the control districts, as this could contaminate the impact measurement if they also affect pupil learning. Although some LANES and Tusome Pamoja activities (mainly training) in control areas seem likely to have contaminated the endline impact estimates to some extent, there are reasons to assume that this contamination is fairly minimal. BRN-Ed training (including a programme called STEP) also affected teachers in treatment and control areas, but again the evidence suggests that this is unlikely to have seriously contaminated the EQUIP-T impact estimates. The details supporting this assumption are provided in Volume II (Chapter 3 Section 3.5.1). There are some implications of these other programmes' activities for the interpretation of the impact estimates, as explained in the next chapter, Section 2.5.1.

2 Impact evaluation objectives, methods and endline design

2.1 Impact evaluation objectives

The overall objectives of the EQUIP-T impact evaluation are to:

- generate evidence on the impact of EQUIP-T on learning outcomes for pupils in primary education, including any differential impacts for girls and boys;
- assess perceptions of the effectiveness of different EQUIP-T components;
- provide evidence on the fiscal affordability of scaling up EQUIP-T; and
- communicate evidence generated by the impact evaluation to policymakers and key education stakeholders.

2.2 Impact evaluation design and methods

At baseline and midline, the evaluation used a mixed methods approach whereby quantitative and qualitative methods were integrated to provide robustness and depth to the research findings. The baseline was carried out in 2014, prior to EQUIP-T's field implementation, and the midline was carried out two years later, in 2016. In both of these rounds the quantitative and qualitative data were collected within a relatively short period of each other, enabling the use of mixed methods both at design, data collection, and analysis stages. The sequencing of the mixed methods had to be changed in the endline design, for reasons explained in the next section, but the two core components of the impact evaluation methodology remain the same, as follows:

1) Quantitative estimation of impact: The design is based on a quasi-experimental approach, with multi-stage sampling. The sample is a panel of 200 schools: 100 programme treatment schools, and 100 control schools, to act as a counterfactual. The map of Tanzania in Figure 1 indicates the treatment districts, the subset of treatment districts that are part of the impact evaluation, and the control districts. The impact identification method combines propensity score matching (PSM) with difference-in-differences (DID)—full details are given in Volume II (Chapter 4).

This analysis produces estimates of the impact of the EQUIP-T programme *as a whole* on various outcome- and impact-level indicators (including pupil learning achievement), but it is not able to quantify the impact of the different EQUIP-T components separately on these indicators. A second method, described below, provides descriptive evidence on the effectiveness of the different components in meeting their objectives, and, under certain assumptions from the programme theory of change, contributing to any programme-level impact.

2) Rigorous factual analysis to explain programme impact¹⁸: This approach combines quantitative survey data on trends in programme areas with qualitative research findings, together with other secondary sources, in order to understand key channels of programme influence, or reasons for lack of change, addressing research questions structured around the programme's theory of change. In other words, this analysis seeks to understand whether, how, and why

¹⁸ The term 'rigorous factual analysis' comes from White (2009).

national priorities (and potential future programming) are among those that have started to be implemented relatively recently.²⁰ These encompass interventions related to empowerment, and tackling cultural norms and taboo issues, which require qualitative research methods. For this reason, among other considerations, it was agreed with DFID that the qualitative endline research will take place in 2019.

The decision to split the endline evaluation into two parts, with the first (quantitative) part carried out in 2018, followed by the qualitative part in 2019, has the clear advantage of addressing the two issues noted above. This had to be set against the drawback of extending the sequencing of the mixed methods across two school years, which means that integrated analysis based on findings collected using different methods at the same time is not possible.

The approach to ‘mixing’ at endline is therefore sequential, with the qualitative analysis serving partly to deepen understanding of selected findings from the quantitative research (Greene *et al.*, 1989), and also to explore priority themes that are not amenable to quantitative methods. As at midline, there will be two levels of endline qualitative research: school/community and regional/district—the latter will provide evaluation evidence on the district planning and management component. There will also be a separate cost study conducted in 2019, following on from the initial analysis done at midline.

2.3.2 Research priorities for the quantitative endline evaluation

The research priorities of this first part of the endline evaluation are captured in the Endline Evaluation Matrix Part I (Annex C). This sets out evaluation questions linked to the programme theory of change, which can be addressed using findings from the three rounds of the quantitative survey. These evaluation questions relate to the expected changes, and assumptions, at each stage along the results chain for the five school- and community-level EQUIP-T components under this first part of the endline evaluation (Table 1).²¹ At impact level, the key question is whether the programme has improved pupil learning achievement. The majority of evaluation questions relate to the teacher performance and SLM components because more of the changes expected to flow from these components can be measured using quantitative methods.

The findings from the thematic (component) analysis, combined with those on pupil learning achievement, feed into overall conclusions that address four of the standard evaluation criteria for development programming²²: impact, effectiveness, relevance, and (to a lesser extent) sustainability.²³ This is necessarily partial and preliminary at this stage, as the findings from the endline qualitative research and costing study are still to come. Efficiency, the fifth standard evaluation criteria, was discussed in the midline report, and will be revisited in the full endline report.

²⁰ This is partly because some original activities have been delayed, and partly due to new activities under the extension.

²¹ 1A Teachers’ capacity and performance, 2A SLM, 2B SIS, 4A Community participation and accountability, and 4B Conducive learning environment marginalised children, particularly girls and children with disabilities.

²² These come from the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC). For definitions of the criteria, see www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm.

²³ Since the programme is ongoing, the evaluation reports on signs of conditions being in place that are necessary for sustainability. See OPM (2017a) for a preliminary analysis of signs of sustainability using the midline evaluation findings.

2.4 Endline survey instruments and actual samples

The endline survey took place in April/May 2018, the same time of year as the baseline and midline surveys in 2014 and 2016, respectively. It is critical that the core information collected over multiple rounds remains the same, so that key indicators can be reliably tracked over time using common questions and metrics. Adaption of the midline instruments for use at endline was thus not substantial, but there were two overarching changes made to the suite of instruments. First, a small group interview with teachers was added, to give information on the delivery of EQUIP-T teacher in-service training at a school level since the start of the programme. Second, the teacher development needs assessment, which measured teachers' subject knowledge, was removed, partly because the final design of EQUIP-T's in-service training did not focus on strengthening subject knowledge.²⁴ More details on the endline instruments (content, respondent, and sampling at school level) are given in Volume II, Chapter 3.

Table 2: Endline quantitative survey sample sizes

Sampling unit	Treatment sample			Control sample		
	Target sample	Actual sample	Actual/Target (%)	Target sample	Actual sample	Actual/Target (%)
Regions	5	5	100	7	7	100
Districts	17	17	100	8	8	100
Schools ¹	100	100	100	100	100	100
Std. 3 pupils (tested both in Kiswahili and maths)	1,500	1,499	99.9	1,500	1,500	100
Parents of tested pupils (poverty scorecards)	1,500	1,495	99.7	1,500	1,497	99.8
Stds. 1–3 Kiswahili/maths teacher interviews ²	441	435	98.6	455	454	99.8
Teachers' group interview on in-service training	100	99	99.0	n.a.	n.a.	n.a.
Std. 2 lesson observation maths ³	100	95	95.0	n.a.	n.a.	n.a.
Std. 2 lesson observation Kiswahili ³	100	101	101	n.a.	n.a.	n.a.

Source: Endline survey. Notes: (1) The school instruments are: head teacher interview and data collection from school records; and head count of teacher and pupil attendance. (2) The samples includes 16 head teachers (treatment) and 25 (control) who teach Kiswahili or maths to Stds. 1–3. All 7 teachers in treatment and control schools who did not sit for the interview were unavailable (absent on the day and could not be reached over the phone later). Some 11% of teachers were interviewed over the phone because they were absent on the day of the survey. (3) 95 maths (arithmetic) lessons and 101 Kiswahili lessons (either reading or writing) were observed. Some of these subjects were taught consecutively (without a break) in one class period. 138 separate class periods were observed.

Response rates to the endline survey were very high. The survey reached all 200 schools, as intended, and Table 2 above shows that actual sample sizes are close to target sample sizes for pupils, their parents, and teachers. The response rate was lowest for maths lesson observations, at 95%. Full details of the evaluation's sampling strategy are given in Volume II of this report (Chapter 3). This also explains why the teachers' group interview and lesson observations have been conducted in treatment schools only.

²⁴ The results from the teacher development needs assessments given at baseline and midline showed little change.

Statistical representativeness and survey weights

Estimates represent pupils or teachers or schools, depending on the indicator, in *programme treatment areas only*. More specifically, the sample is representative of Government primary schools in 17 programme treatment districts drawn from the five regions in which the EQUIP-T programme started implementation in August 2014 (Figure 1). These districts are fairly similar, in terms of key contextual factors, to the 12 other districts in the early EQUIP-T programme, but overall the EQUIP-T districts are significantly more socially and economically disadvantaged than the remaining districts in Tanzania (see OPM (2015a), p. 8). This means that the descriptive quantitative estimates in this report represent estimates for disadvantaged districts and should not be generalised to the national level.

Having said this, the agreed TOR for this impact evaluation study (Annex A) explain that: ‘a large majority of rural districts in Tanzania share similar characteristics to those in the EQUIP-T treatment sample and therefore although the results will not be statistically generalisable outside the impact evaluation sample, it is reasonable to expect that the findings will have some applicability in other districts as well, if sufficiently similar to the treatment districts’.

All estimates are adjusted using survey weights to account for the sampling strategy. The teacher estimates are based on school-level populations—all teachers, or all teachers of Standards 1 to 3 who teach Kiswahili or maths. The survey weights have been adjusted to account for this. Lessons were sampled by convenience.²⁵

Some tables and figures compare estimates for different sub-populations (for example, boys and girls) at the same point in time, but it is important to note that no causal inference can be made from these simple disaggregations because they are just descriptive. Due to sample size limitations, it is not possible to disaggregate estimates by region or district.

2.5 How to read the report

2.5.1 Impact estimates

As discussed in Section 1.4.2 in Chapter 1, two other early grade learning programmes, LANES and Tusome Pamoja, are likely to have contaminated the EQUIP-T impact estimates to some extent because of their operations in the impact evaluation control districts. While LANES is a national programme, some of its earlier activities were not national in coverage and were carried out in non-EQUIP-T districts only including the evaluation control districts.²⁶ Tusome Pamoja operates in four mainland regions, including one of the evaluation control districts (see Volume II Annex H for more details on the implementation of both programmes). This has implications for the interpretation of the impact estimates presented in this report. Specifically, the impact identified by the analysis represents the effect that EQUIP-T as a package has had on outcome indicators, compared to a counterfactual situation where, in the same schools, the alternative Tusome Pamoja and non-national LANES training and materials have been delivered. In other words, the analysis **measures the added impact of all EQUIP-T-related interventions over and above the potential effect of the non-national LANES initiatives and Tusome Pamoja initiatives.** If

²⁵ This means that enumerators observed the available Standard 2 Kiswahili or maths lessons that were operating on the day of the survey. Typically the first available lessons were observed, but sometimes this was not possible due to teacher absence or sometimes due to other survey activities.

²⁶ Most LANES activities are national in coverage, affecting both treatment and control schools, and these in principle should not affect the impact estimates. The contamination comes from non-national LANES activities that took place in the evaluation control schools but not the treatment schools. For example the Standards 1 and 2 curriculum (3Rs) orientation training that took place in 2015 did not include teachers from EQUIP-T regions.

certain assumptions hold (see Box 2), including those related to the BRN-Ed/STEP training then the full impact of EQUIP-T may in reality be slightly higher than the estimates in this study.

Box 2: Assumptions related to the interpretation of the impact estimates

Although there are reasons to assume that the extent of contamination from LANES and Tusome Pamoja in the evaluation's control districts is fairly minimal (see Volume II, Chapter 3, Section 3.5.1), it is also reasonable to assume that if these interventions have had any effect on the outcomes being measured in this study (such as pupil literacy and numeracy levels), it is likely to have been positive. Under this assumption, outcome levels in the control group schools are, on average, higher than they would have been in a pure counterfactual situation (with no contamination). This in turn means that the impact of EQUIP-T, which is estimated by comparing treatment schools and control schools over time, is potentially a slight underestimation of the full impact of EQUIP-T as a whole. This ignores the possibility that the effects of the BRN-Ed/STEP training which took place to a limited extent in both treatment and control areas, had a greater positive effect on pupil learning in treatment areas than in control areas. If in fact the latter is true, then this would reduce any potential overestimation of EQUIP-T impact due to LANES or Tusome Pamoja contamination of the control areas.

In the present volume, impact estimates are presented in boxes with black borders, to distinguish them from descriptive estimates of trends in outcomes in programme schools. Each impact box contains a graph and explanation: see Box 6 in Chapter 3 for an example. Each graph shows point estimates for the average treatment effect on the treated (ATT) on outcome indicators and 95% confidence intervals for these effects. This means that the probability that the true treatment estimate will fall within this area is 95%. The ATT is the average effect that the treatment (EQUIP-T) had on pupils or teachers in the EQUIP-T schools (the treatment group).²⁷

Outcome indicators used in this evaluation are mainly proportions. Estimates of treatment effects are given in percentage point changes for these proportions. For example, if the estimated ATT on the proportion of pupils in the top performance band for maths in treatment schools is 0.03, this means that EQUIP-T has increased this proportion by three percentage points (for example, from 10% to 13%), compared to a counterfactual situation. Kiswahili and maths test score results (scaled as Rasch scores, explained in Chapter 3) are the only outcome indicators analysed that are not proportions. In those cases, the estimates of treatment effects are given in logits and also expressed in standard deviation (SD) units.²⁸ When confidence intervals of such estimates do not overlap with zero, this is an indication that this treatment effect is truly different from zero. A zero value is indicated using a red line in the graphs. Text under each graph explains the level of robustness and confidence in these findings, including information from the main estimation model and robustness checks. The results are presented in detail in Volume II (Chapter 4).

Impact estimates are found in two places in this present volume. Estimates of the impact of the programme on pupil learning outcomes are in Chapter 3 in Box 6, while estimates of programme impact on teacher absenteeism are in Chapter 4 in Box 9 (reference is also made in Chapter 5 to the estimate of programme impact on head teachers' use of performance appraisal which is detailed in Volume II, Chapter 5). All other charts and tables which display survey results contain estimates of descriptive trends in programme areas (see details below), not estimates of programme impact.

²⁷ It is useful to distinguish between ATE (Average Treatment Effect) and ATT (Average Treatment Effect on the Treated). The latter (ATT) measures the causal effect of the treatment for individuals in the treatment group only; the former (ATE) measures the causal effect of the treatment across all individuals in the population from which the treatment group comes from. With the PSM and DID estimation approach in this evaluation, ATT is measured.

²⁸ The basis of the SD units are the distributions of endline pupil learning scale scores for Kiswahili and for maths.

2.5.2 Descriptive trends in programme areas

Most of the quantitative evidence presented in the findings chapters which follow takes the form of descriptive trends in **key indicators in programme schools** between baseline (BL), midline (ML), and endline (EL). The tables typically have the following headings, and include the term ‘trends in programme areas’ in the title:

Indicator	Baseline estimate	N	Midline estimate	N	Endline estimate	N	Difference BL-EL	Difference ML-EL
-----------	-------------------	---	------------------	---	------------------	---	------------------	------------------

For each indicator, the baseline, midline, and endline estimates are shown together with their sample sizes (N). The columns marked ‘Difference’ contain estimates of the change in the value of the indicators from baseline to endline, and from midline to endline. If this change is statistically significant based on simple t-tests, this is marked with asterisks (*significant at the 10% level, **significant at the 5% level, ***significant at the 1% level). The more asterisks that are shown, the more likely it is that the observed change is due to a real change over time, rather than due to chance related to who was interviewed or tested. Where results are not given an asterisk this does not mean that there is no change over time, but rather that the difference cannot be asserted with such a high degree of confidence (90% certainty or more that the change is not due to ‘chance’). For completeness, the statistical tables in Volume II (Annex G) also contain the t-tests for the baseline to midline differences in indicators.

Note that this report does not contain descriptive trends for the control group of schools because directly comparing results from programme (treatment) and control schools, teachers and pupils would be misleading. The quasi-experimental design of the evaluation means that the control schools were sampled in such a way that they are not representative of any underlying population. The only way in which treatment and control schools, teachers and pupils are comparable is through the impact estimation approach (which combines PSM with DID, and controls for selection bias). This is explained in more detail in Volume II Chapter 3 Section 3.4.2.

2.5.3 Red, amber, green (RAG) summary results tables

The endline survey has produced a vast number of findings, which are set out in the five results chapters which follow. To help the reader to digest the main findings, a summary table of evidence is provided at the end of each of the component chapters (Chapters 4 to 7). This pulls out the main indicators from the chapter, and colour codes the results using the simple classification in Figure 2. This is simply meant as a communication device that uses a classification that is easy to remember. Of course, the choice of indicators is somewhat subjective, but the aim is to give a visual snapshot of the extent to which the quantitative evidence shows that the component is achieving its objectives.

Figure 2: Classification of indicator values in summary results tables

RAG classification of indicator values	
Red	< 50%
Amber	>=50% and <75%
Green	>=75% and <90%
Dark green	90% or above

Source: OPM

2.5.4 Terminology

The following terms are used repeatedly and have particular definitions in this report:

Box 3 Key terminology used in this report	
Term	Meaning in this report
Impact	A causal effect.
Kiswahili skills	This is used synonymously with 'literacy'. For example, when findings refer to pupils' Kiswahili skills, it means literacy skills (reading and writing) in the Kiswahili language.
LGAs	This is used synonymously with 'districts'.
Maths skills	This is used synonymously with 'numeracy'. For example, when findings refer to pupils' maths skills, it means numeracy skills (arithmetic) as part of maths.
Programme schools	This is used synonymously with 'EQUIP-T schools' and with 'treatment' schools.
PTPs	PTPs are school-level bodies, established as part of EQUIP-T to support schools, comprising parents and teachers. They do not have formal governance authority.
Significant	Statistically significant.
Staff Turnover	The rate at which staff leave their current posts that are defined by their role (e.g. teacher or head teacher) and place of work. Therefore teachers who transfer between schools are included in the turnover rate, as are head teachers who are demoted within the same school because both have left their current posts.

Source: OPM.

2.6 Structure of Volume I

This is Volume I of the EQUIP-T endline quantitative evaluation report, which is accompanied by *Volume II: Methods and Supplementary Evidence*. The remainder of this volume is structured as follows:

Part B contains the endline quantitative findings. There are five chapters: the first (Chapter 3) covers pupil learning and includes estimates of the impact of the programme on learning achievement in Kiswahili and maths. The next four chapters correspond to the four EQUIP-T sub-components covered in this study: teacher performance (Chapter 4); SLM (Chapter 5); community participation and accountability (Chapter 6); and conducive learning environment for girls and marginalised children (Chapter 7). Each of these component chapters is structured in the same way, as follows:

- There is an introduction, which provides a diagram of the relevant results chain from the theory of change; details of programme implementation up to the endline survey; and a description of expectations of change (according to the EQUIP-T logframe targets, where available, and interviews with programme staff).
- The findings section is then structured using the results chain from the theory of change. Evidence is presented on the extent to which inputs have been delivered, and whether changes expected to flow from these inputs have happened in the programme areas as set out in the theory of change. The findings section also assesses whether the theory of change assumptions have held—these reflect the optimal conditions needed to support changes – for example, low teacher turnover or appropriate class sizes.
- There is then a short summary of evidence section, which includes the RAG table described above.

Part C (Chapter 8) discusses the implications of the findings from Part B in terms of programme impact, effectiveness, relevance, and sustainability, and makes recommendations for programme consolidation and adjustment, as well as drawing wider lessons.

Part B: Findings

MAANDALIO YA SOMO (TEACHER'S LESSON PLAN)

(IDADI YA WANAFUNZI) (NUMBER OF PUPILS)

TARUHU (DATE)	SOMO (SUBJECT)	DARASA (CLASS)	VIFUNZI (PERIOD)	SUDA (TIME)	WALIOJANDIKISHWA (REGISTERED)			WALIOJIKUNDA (ATTENDED)	
					WAV	WAS	JML	WAV	WAS
14/1/20	KISWA	II	2	6 ⁰⁰ -6 ³⁰					

UJUZI / COMPETENCE: KUOLEWA NA KUTUMIA LUIGI

LENGO KUU / MAIN OBJECT: Wanafunzi watahamu mwanalimu

MADA KUU / MAIN TOPIC: MWANDIMU

MADA NDOGO / SUB TOPIC: Kusoma

MALENGO MANUSUJI / SPECIFIC OBJECTIVES: Biada ya kipindi mwanalimu
auzi: kusoma.

ZANA / VIPAA / TEACHING AIDS: Chati, vitabu

KEJEA / REFERENCE: Kitabu cha mwanafunzi

MCHAKATO WA UFUNDISHAJI (TEACHING PROCESS)

HATUA (STAGES)	SUDA (TIME)	VITENGO VYA UFUNDISHAJI (TEACHING ACTIVITIES)	VITENGO VYA UJIFUNZAJI (LEARNING ACTIVITIES)	VITENGO VYA UHAKAMA (ASSESSMENT)
1. UTANGULIZI (INTRODUCTION)	5	- Kuwongeza kwa kutoka ma- lizo mafupi.	- Kumbuliza	- Anasikia
2. UJUZI MPYA (NEW KNOWLEDGE)	10	- Kuwongeza ma- nu mwanafunzi mwanafunzi	- Kusoma	- Anasikia
3. KUIMARISHA MAARIFA (REINFORCEMENT)	3	- Kuwongeza ku- kumbuliza ma- suti.	- Kupitua maswali	- Wanaweka ku
4. KUTAPAKARI (REFLECTION)	2	- Kuwongeza wendele kuoma	- Kuendelea kusoma	- Wanaweka kusoma
5. HTIMISHO (CONSOLIDATION)	10	- Kuwongeza wendele kusoma	- Kuendelea kusoma	- Wanaweka kusoma

TATHIBI YA MWANAFUNZI: _____

PUPU'S EVALUATION: _____

TATHIBI YA MWANALIMU: _____

TEACHER'S EVALUATION: _____

MADNI /REMARKS: _____

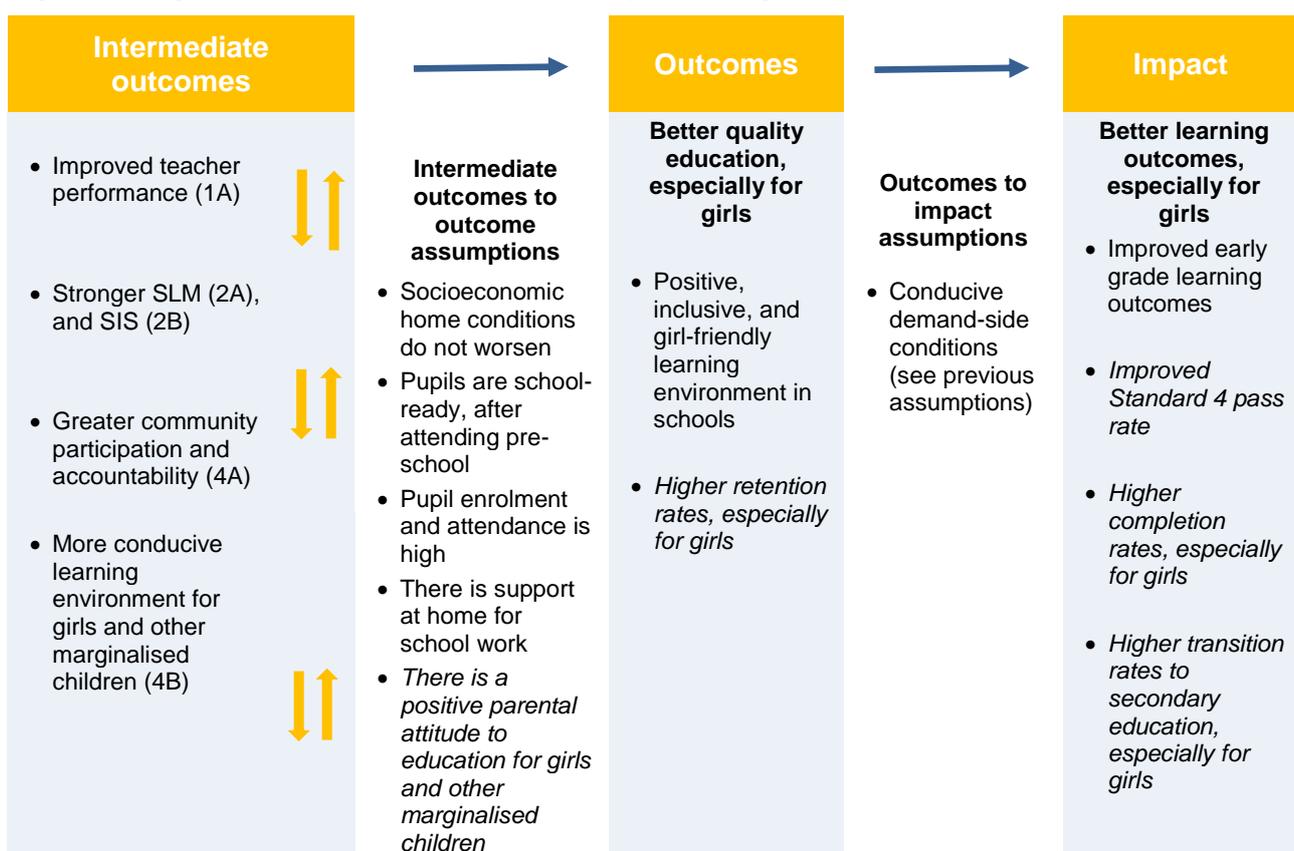
3 Pupil learning, background, and home support

3.1 Overall programme goals and expectations of change in pupil learning by endline

The overall goal of EQUIP-T is to achieve better learning outcomes at primary level, especially for girls. Figure 3 shows how the results (intermediate outcomes) from the five programme components covered in this quantitative part of the endline evaluation are intended to contribute to the programme's desired outcomes (better quality education, especially for girls, and higher retention) and impact (better learning outcomes, especially for girls, for an increasing number of pupils). The arrows beside the components indicate that they are intended to be mutually reinforcing in leading to better quality education, and to result in greater change than if they were implemented in isolation.

The main assumptions underpinning the high-level theory of change are that key demand-side conditions, such as regular pupil attendance, do not worsen and do not pose overwhelming barriers that reduce, or negate, the effects of the programme interventions. In some cases the programme is actively seeking to influence these conditions – for example, in improving access to pre-primary education.

Figure 3: High-level results chain from theory of change^{1, 2}



Source: OPM (2018) (Endline evaluation matrix (partial), see Annex C). Notes: (1) This is a partial high-level results chain, which covers programme components covered by the quantitative endline evaluation. It also excludes the institutional strengthening aspect of the theory of change, whereby the programme seeks to support national adoption of successful aspects of the EQUIP-T programme. (2) The indicators and assumptions in italics are not being measured in the impact evaluation quantitative survey.

Below this high-level results chain sit the expected results chains for each of the programme components, running from inputs to outputs to intermediate outcomes. The four chapters which follow the current chapter present quantitative evidence related to these component results chains. As noted in Chapter 2 (Section 2.3) the quantitative evidence is more comprehensive for the teacher performance and SLM components, than for the others.²⁹ It is important to bear this in mind, when reading this chapter, which draws findings from the component analyses to help explain programme impact.

The main objective of this chapter is to answer the key evaluation question of whether EQUIP-T has had an impact on early grade learning achievement, after four years of implementation. It also seeks to explain the most likely reasons for any impact, drawing on descriptive evidence related to the theory of change that is discussed in detail in the component chapters (Chapters 4 to 7). The other key results relate to trends in overall levels of early grade learning in the programme regions, and changes (if any) in learning disparities by gender, poverty status of the pupils' household, and main language spoken at home. This is supplemented by results discussed in Volume II Chapter 5 covering trends in learning outcomes in programme schools for various sub-populations of EQUIP-T beneficiaries (girls, boys, pupils from poorer and richer homes, pupils whose home language is or is not Kiswahili) over the four year period. It also includes a description of changes in learning outcomes between midline and endline for pupils in programme schools with disabilities. Given EQUIP-T's increasing focus on inclusive education in the extension period up to January 2020, these results should be of particular interest to the MA.

To give some context to the results, Box 4 contains information on the scale of expected change in learning achievement in the programme regions over the programme lifetime.

Box 4: Learning achievement targets from programme logframes, and original intervention summary

Source 1: Original Intervention Summary for EQUIP-T (DFID, 2013, p. 38)

Indicator: proportion of early standard pupils with basic literacy skills

Target: to more than double the value of this indicator over the original programme period (2014 to 2018)

Source 2: EQUIP-T logframes (EQUIP-T MA, various years)¹

Indicator: proportion of Standard 3 pupils achieving Standard 2 literacy and numeracy skills in the five original programme regions (those covered by the impact evaluation)

Baseline: 12% literacy, 4% numeracy

Target (2018): 21% literacy and 10% numeracy, with no gender disparity

Target (2019): 25% literacy and 10% numeracy, with no gender disparity

Note (1): The target for 2019 comes from the latest logframe made available to the evaluation team (dated November 2017), but as there is no target for 2018, this was taken from an earlier logframe (November 2015).

This chapter is structured into six further sections, starting with Section 3.2 which presents trends in pupil background conditions since baseline, followed by trends in early grade learning results in Section 3.3. This section also contains the impact estimates, which demonstrate whether part of the gains in learning achievement can be robustly attributed to the EQUIP-T programme or not. All impact estimates in this report (which are based on a comparison of EQUIP-T programme schools against a counterfactual set of control schools) are presented in shaded boxes, in the interests of clarity. These key results are followed by a brief summary of the most likely reasons for any observed impact based on the main findings from the subsequent EQUIP-T component chapters (Chapters 4 to 7).

²⁹ This is particularly true for Component 3A, district planning and management, where only the WEOs' support to school is assessed, and hence this report omits the majority of the interventions under this component.

Sections 3.4 to 3.6 discuss changes in learning gaps by gender, home language, and household poverty in the programme schools since baseline. The discussion of these descriptive trends is complemented by summary evidence from the component chapters which point to likely reasons for some of the changes observed. The final section, 3.7, summarises the key points regarding changes in pupil learning.

3.2 Trends in pupil backgrounds and home support for education in treatment areas

As highlighted in Figure 3 above, pupil backgrounds and home support factors can inhibit (or enhance) the link in the theory of change between improving schooling and achieving the desired learning gains.

Before presenting trends in learning achievement of Standard 3 pupils since baseline, it is first useful to see if there are any changes in the profile of Standard 3 pupils over time, as these could be contributing to changes in learning levels. Changes in pupils' profile cover personal characteristics, such as gender and age; economic circumstances; support for education at home; and the demand for schooling. It is worth noting that the majority of the current cohort of Standard 3 pupils in 2018 entered school in Standard 1 in 2016, shortly after the fee-free basic education announcement, so it might be expected that the profile of this cohort differs from previous cohorts.

EQUIP-T is seeking to directly influence some of the background factors discussed below, mainly via its community participation and inclusive education components (4A and 4B, respectively) (discussed in Chapters 6 and 7). In addition, EQUIP-T piloted, and has subsequently expanded, a community-based short pre-school programme, called the SRP. This is now part of component 1B (improving rural access to quality pre- and early primary) under the programme extension. Although this evaluation does not cover this element, Box 5 below explains more about the intervention and summarises findings from EQUIP-T's commissioned research.

3.2.1 Has the personal and economic profile of Standard 3 pupils changed since baseline?

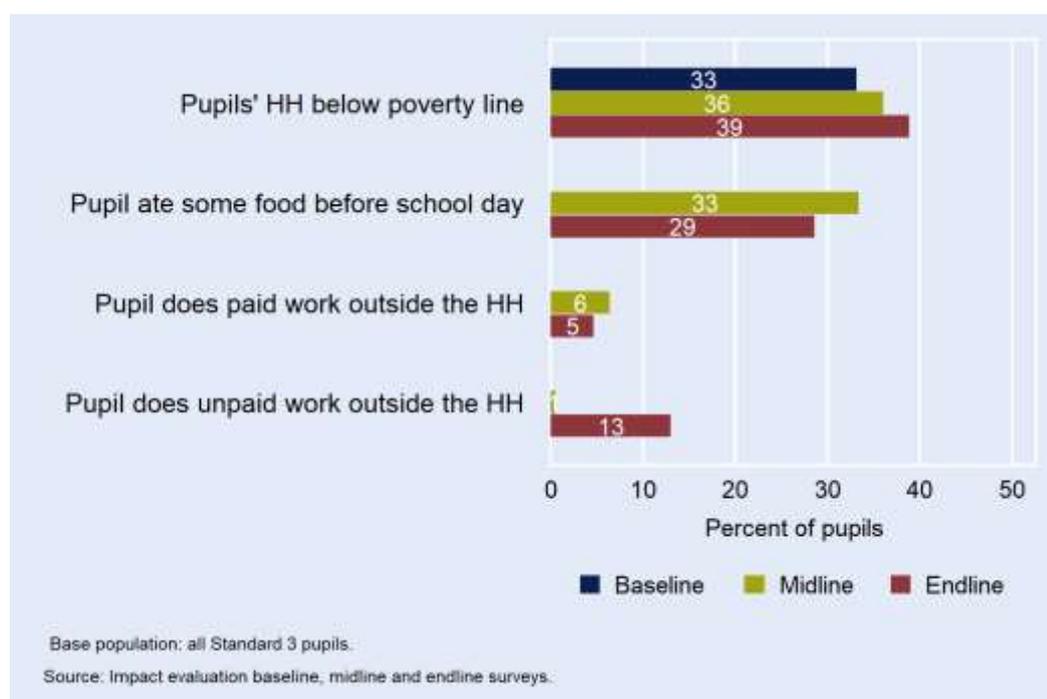
The near equal balance between boys and girls attending school in Standard 3 (52% girls at endline), as well as their average age of 10 years, has not changed significantly since baseline. The distribution of pupils' ages has changed significantly, however, such that the share of **over-age pupils (those greater than 11 years) increased from 36% to 46%**. This appears to be due to a greater share of over-age pupils entering school for the first time three years earlier, shortly after the fee-free policy announcement, rather than due to high repetition rates. Only 1% of Standard 3 pupils report that they are repeating the standard. Over-age enrolment is problematic for a number of reasons: it can affect demand for schooling (premature dropout), as well as the effective delivery of education and the efficiency of the system.³⁰ One of EQUIP-T's logframe targets is to reduce the share of over-age children entering Standard 1 in its programme regions from 20% in 2014 to 12% by 2018. Of course, this may be a temporary problem if the fee-free policy caused a one-off surge in older children returning to school or entering for the first time.

Pupils' economic circumstances have worsened to some extent since baseline. Close to 40% of Standard 3 pupils come from households that are predicted to fall below the national

³⁰ See, for example, evidence from studies which find that over-age pupils are more likely to repeat grades and drop out of school before completing a cycle (UN Economic, Scientific and Cultural Organization (UNESCO), 2007).

poverty line,³¹ up from 33% at baseline (Figure 4), and this change is weakly significant. The programme deliberately selected remote and economically disadvantaged districts into the programme, and so this extremely high share of pupils coming from low socioeconomic backgrounds was expected from the outset. Figure 4 also illustrates how the worsening economic circumstances affect pupils directly. Particularly striking is the **large and strongly significant increase in the share of pupils doing unpaid work outside the household**, which reached 13% by endline.³² There is also a drop of five percentage points in the already low share of pupils who ate some food before school, on the day of the survey, to under 30%, and this change is weakly significant. Being hungry at school can affect the ability of pupils to concentrate on learning (Pivak et al, 2012, and Hoyland et al, 2009), and about 10% of schools have a feeding programme to try to mitigate this (see Chapter 7).

Figure 4: Trends in pupils' economic background (trends in programme areas)



3.2.2 Has the home support for education and demand for schooling for Standard 3 pupils changed since baseline?

The vast majority of pupils (81% at endline) do not speak Kiswahili, the language of instruction, as their main language at home, and this share has not changed significantly since baseline.³³ While this is an important environmental factor, and a marker of disparity in learning achievement (see Section 3.5), there are some signs of positive change in home support for education. Just under half of pupils' homes have books, newspapers, or other reading materials available at endline, a significant and large gain of nine percentage points since midline. Another positive change is in the availability of someone at home to help pupils with homework. Some 70% of pupils have support for homework now, significantly up from 63% at midline. Direct support for reading, however, is unchanged since midline. A sizeable minority of pupils *never* read out loud to

³¹ See Volume II, Chapter 3 for further details on how this was measured.

³² The survey did not probe for examples. In rural Tanzania there is a communal culture of helping each other with activities such as collecting water, firewood, taking care of younger children, and grazing cattle.

³³ This should not be taken to mean that this group of pupils do not understand Kiswahili, but that they have less opportunity to practise speaking Kiswahili than their peers who come from homes where Kiswahili is the main language.

someone at home (about 20%) or *never* have someone at home read out loud to them (about 35%).

Nearly three-quarters of Standard 3 pupils say they attended pre-school, and there is no significant change in the reported rate since midline.³⁴ Of those with pre-school experience, about 80% attended a Government pre-school, while 20% attended a nursery or kindergarten. Only one pupil reported attending an EQUIP-T SRP, which most likely would have been when the SRP was being piloted in 2015 (see Box 5 for more details). Both at midline and endline, parents reported that 5% of Standard 3 pupils attend extra paid classes, so having this extra support outside the home is fairly rare.

Box 5: EQUIP-T's SRP

SRP implementation summary

The SRP started in October 2015 with a 12-week programme in 25% of programme districts. This included 1,000 centres and 50,000 enrolled children. The SRP is based in community buildings and is delivered by community teaching assistants (now called SRP facilitators), who are trained in a five-day programme prior to teaching starting. To support this, the programme developed an SRP competency framework to guide learning (focusing on building competencies that support rapid learning), including 12 SRP story-books. The framework is based on active pedagogy and a story-based approach. SRP facilitators further learn in training to make and use locally produced teaching aids. The number of SRP centres expanded to 2,744 in October 2016, when another 12-week programme was delivered. In 2017, the duration of the SRP was expanded to 16 weeks.

Three main aims of SRP

Access: To provide some pre-schooling to children who are out of school and likely will not attend pre-school at all.

Equity: To target areas where there is no existing pre-school (more remote and typically more disadvantaged areas) and where children do not speak Kiswahili (since these children perform much worse in the early grades of primary school and their need for school readiness interventions is greater). The programme has a focus on inclusive pedagogy, particularly with regard to addressing gender imbalance.

Quality: By introducing an active pedagogy using a story-based approach, the theory is that this will enable the children to get prepared for school faster (i.e. this pedagogy will be more effective and more engaging for young children).

Findings on SRP provision from the impact evaluation quantitative surveys

Between the midline and endline surveys, the share of head teachers in programme schools reporting that there is an SRP centre in the community increased significantly from 18% to 51%, in line with the scale-up from a pilot programme. This highlights both impressive growth in SRP centres over a short period, but also the limited coverage of the SRP, which is not present in just under half of the communities serving programme schools.

Findings from EQUIP-T's SRP research study

Results from an international development and early learning assessment (IDELA)¹ conducted in March 2017 in the seven EQUIP-T regions, found that Standard 1 pupils who had attended an SRP had significantly stronger skills than those who had no access to early childhood education, and moderately stronger skills than those who had attended formal pre-school. These findings are based on a sample of 1,119 Standard 1 pupils from 42 schools in 14 districts across the seven regions. Within each school, pupils were grouped according to their pre-school experience, and then sampled randomly. The study does not control for background characteristics of pupils, so the differences observed in performance cannot be interpreted as being caused by the different types of pre-school exposure.

Source: EQUIP-T annual monitoring reports (various years), and interviews with EQUIP-T programme staff in January 2018. EQUIP-T MA (2017b). Note: (1) IDELA was developed by Save the Children and contextualised with technical assistance from the Tanzania Institute of Education, the University of Dodoma, and Aga Khan.

³⁴ This was not collected at baseline. At endline (only), parents were also asked whether their child had attended pre-school. The parental reported rate of pre-school attendance for Standard 3 pupils was slightly higher, at 78%.

As an indicator of the **vastly increased demand for schooling since baseline**, Table 3 shows that **mean enrolment in Standard 3 grew from 65 to 95 pupils from baseline to endline**, a strongly significant change of nearly 50%. At the same time, mean enrolment across all standards expanded by 111 pupils on average, a 23% expansion, demonstrating that growth is not confined to Standard 3 (the cohort that entered school shortly after the fee-free education policy). High enrolment growth puts pressure on the schooling system, and the near parallel growth in Standard 3 enrolment and class sizes over the period suggests that the system has not been able to increase basic resources (teachers and classrooms) to cope. The mean class size in Standard 3 at endline is 88 pupils.

Parents are sending their children to school somewhat more frequently than at baseline, but pupil absenteeism is still high. For pupils in Standards 1 to 3, absenteeism fell significantly from 34% to 28% between baseline and endline, and there has been a slightly larger decline for girls than boys (Table 3).

Table 3: Trends in demand for schooling, supply response, and pupil absenteeism (trends in programme areas)

	Baseline		Midline		Endline		Difference	
	Estimate	N	Estimate	N	Estimate	N	BL-EL	ML-EL
Enrolment and class sizes (mean)								
No. of pupils in Stds. 1–7 per school	486.42	100	472.35	100	597.11	100	110.69***	124.76***
No. of pupils in Std. 3 per school	65.18	99	63.87	99	94.82	99	29.63***	30.95***
No. of Std. 3 pupils per class	59.86	99	61.92	99	88.17	99	28.31***	26.26***
Pupil absenteeism on the day of the survey from head count (mean % Stds. 1–3 pupils)								
All pupils	33.50	97	25.17	97	27.94	97	-5.56**	2.77
Boys	34.61	97	26.09	97	29.77	97	-4.84*	3.68**
Girls	32.39	97	24.39	97	25.92	97	-6.47**	1.53
Sources: Evaluation baseline, midline, and endline surveys (school records and head count instruments).								
Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.								

3.2.3 Summary of trends in the profile of Standard 3 pupils

Compared with baseline, the endline cohort of Standard 3 pupils are more likely to be over-age; their economic circumstances are worse to some extent; and they are much more likely to be doing unpaid work outside the household on top of their schooling. On the other hand, there is some positive change in support for education at home, and there is vastly increased demand for education, although this has led to very large class sizes. At the same time, pupil absenteeism has fallen slightly since baseline.

3.3 Trends in early grade learning and impact of EQUIP-T

The purpose of testing Standard 3 pupils is to measure changes in learning achievement over time, as a means of assessing the overall impact of EQUIP-T in relation to its goal. As in previous rounds of the evaluation, the estimates of pupil learning achievement in Kiswahili and maths are measured on scales that are directly linked to early grade curriculum competencies. This provides insight into whether pupils are performing at, above, or below the curriculum level expected, and gives information on the skills different groups of pupils have (Volume II, Annex E, contains more

details on the method of Rasch analysis used to estimate the scale scores³⁵). The current chapter also presents selected ‘raw’ score results, including Kiswahili reading speed, because similar indicators are used to monitor other large-scale programmes in Tanzania. An analysis of trends in raw score pupil learning indicators is given in Volume II, Chapter 5.

The evaluation uses the same adapted-Early Grade Reading Assessment (EGRA) and adapted-Early Grade Mathematics Assessment (EGMA) instruments as at baseline and midline, to test Standard 3 pupils on Standards 1- and 2-level skills. Although the Standards 1 and 2 curricula changed between the baseline and midline, the instruments are competency-based and, with some assumptions, it is reasonable to consider them valid for comparative purposes (see Volume II, Annex E, for further explanation).

3.3.1 How has early grade pupil learning changed overall in treatment areas?

Consistent with the approach taken at baseline and midline, pupils have been estimated to be achieving at the level of one of five curriculum-linked performance bands: **Band 0: below Standard 1 level; Band 1E: emerging Standard 1 level; Band 1A: achieving Standard 1 level; Band 2E: emerging Standard 2 level; and Band 2A achieving Standard 2 level**. Pupils who are ‘achieving at band level’ are more likely than not to be able to demonstrate the skills required for the relevant performance band. A list of competencies linked to each band is given in Volume II, Annex E.^{36 37}

3.3.1.1 Kiswahili results in treatment areas

Pupils’ literacy skills in Kiswahili have significantly improved on average between baseline and endline, and the improvement is particularly strong for the poorest performing pupils.

The bulk of the skills improvement happened between baseline and midline, with a strongly significant rise in average scale scores over the first two years of the programme of 0.9 logits (0.8 SD). There was a further gain between midline and endline, but it is smaller at 0.2 logits (0.2 SD) and only weakly significant.

Figure 5 illustrates the dramatic positive shift in the distribution of pupils across the performance bands, between baseline and endline. It is clear that pupils have, on average, moved up the performance bands over the four years and there is a strongly significant decline in the share of pupils falling into the bottom performance band from 39% to 16%. There is also significant change at the top of the distribution, with the proportion of pupils performing at either emerging Standard 2 level or achieving Standard 2 level rising from 36% to 50% over the period. By endline the share of pupils achieving in the top band (18%) is close to the logframe target for 2018 of 21%.

Put in terms of the absolute number of pupils achieving at each level, the change over time is even more marked because of the sharp growth in Standard 3 enrolment between baseline and endline. Estimates of the number of pupils in government primary schools achieving at Standard 2 level Kiswahili in the 17 districts represented by the evaluation survey sample, reveal an increase from

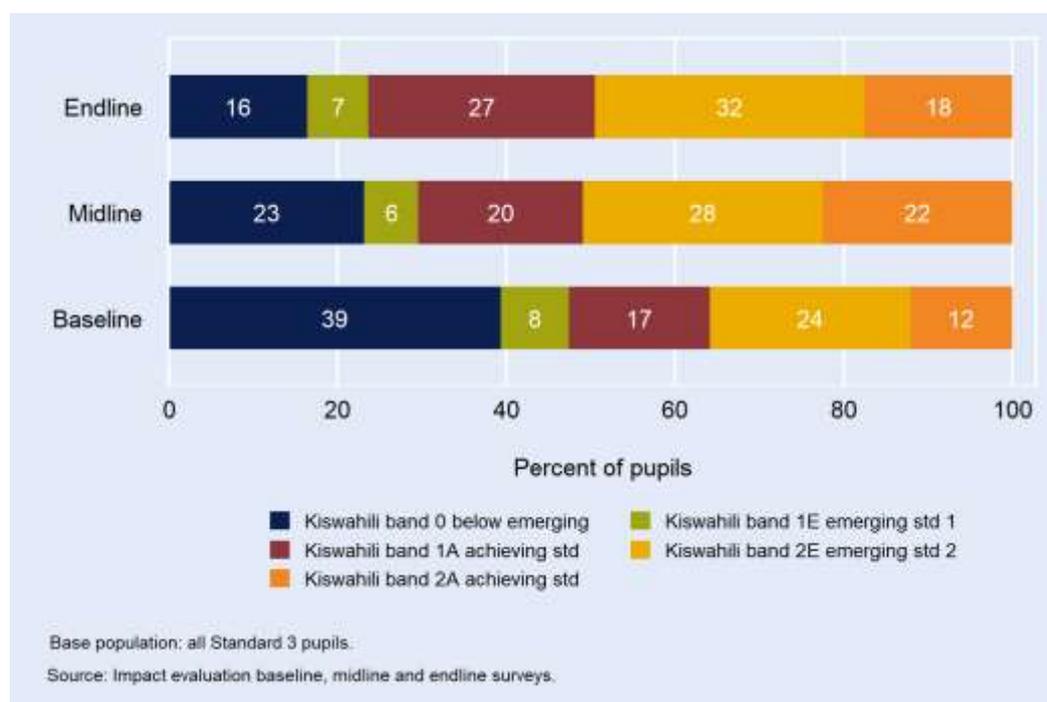
³⁵ Rasch analysis uses a probabilistic model of item response that generates estimates of pupil scores on an interval scale that capture more precisely differences in performance by weighting items by difficulty, compared to a simple raw score analysis.

³⁶ For a minority of items (questions) in the tests and hence the related competencies, there have been changes to their estimated location on the scale over time, these are explained in Volume II, Annex E.

³⁷ The band boundaries scores are the same as at the baseline, for reasons which are explained in Volume II, Annex E.

16,510 pupils at baseline to 35,596 pupils at endline—more than a two-fold increase over the four years. Chapter 5 in Volume II contains the details of this analysis.

Figure 5: Distribution of pupil learning in Kiswahili by performance band in treatment areas, baseline, midline, and endline (%) (trends in programme areas)



While the gain in early grade Kiswahili skills for pupils over four years is remarkable, it is important not to lose sight of the extent to which pupils are still behind curriculum expectations. About half of Standard 3 pupils are achieving at Standard 1 level or below, and are thus at least one year behind in skills acquisition, and therefore need further support to catch up.

Achieving sufficient reading fluency for reading comprehension is one of the key differences between Standard 2 and Standard 1 curriculum standards. The Government's benchmark for Standard 2 pupils is to read at a speed of 50 words per minute (in line with international research on the minimum rate needed for comprehension (Abadzi, 2006)).³⁸ Pupils in programme schools read a simple story passage at a rate of 21 words per minute on average at baseline. This increased to 30 words per minute at midline, but then did not change significantly in the two years between midline and endline (see Volume II, Chapter 5). While the overall gain between baseline and endline is significant and sizeable, it is still far short of the benchmark.

That the gap between current achievement and curricular expectations remains large for such a high proportion of pupils raises the question of whether the expected achievement levels at each Standard may themselves be contributing to poorer performance. Pritchett and Beatty (2012) argue, using data from three studies in South Asia and Africa, that low learning gains over time are partly a result of curricular paces moving faster than the pace of learning (children who are left behind find it increasingly difficult to catch up as the curricula moves on), and that by slowing curricular delivery, cumulative learning can increase.³⁹

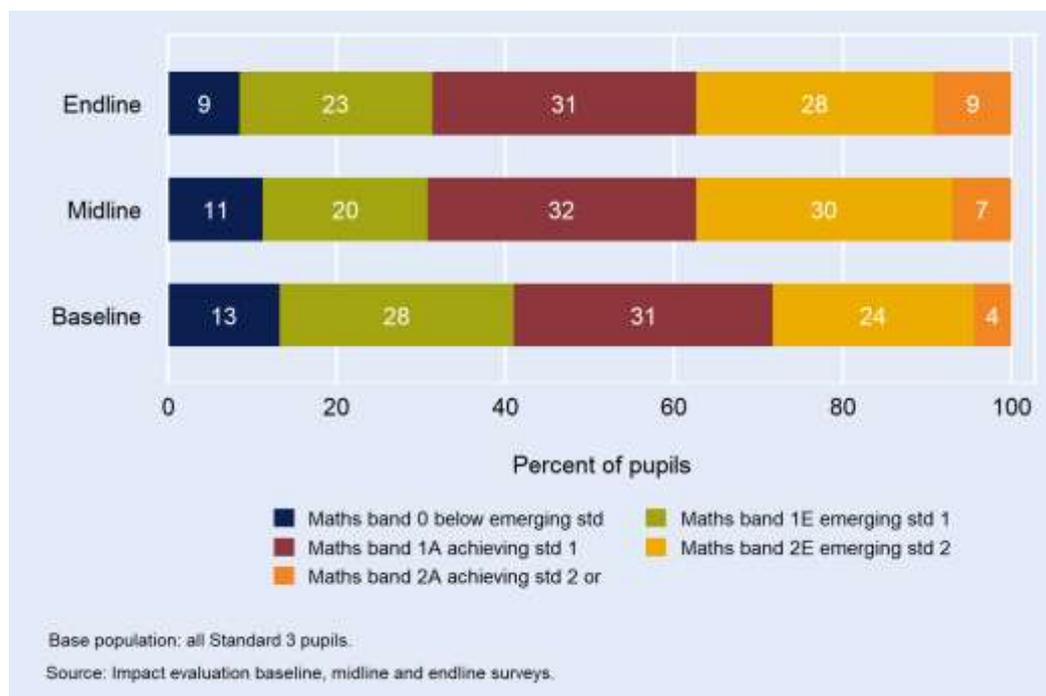
³⁸ MOEST and PO-RALG, 2018b, p153.

³⁹ The paper's findings are based on a simulation model of learning related to the gap between pupil skills and instructional level. The results of the simulations also replicate existing experimental findings.

3.3.1.2 Maths results in treatment areas

Pupils' numeracy skills significantly improved on average between baseline and endline, but the gain is not as marked as for Kiswahili skills. The significant gain in average scale scores for maths between baseline and midline of 0.4 logits (0.2 SD) has been maintained to endline, but there was no additional significant gain in the last two years.

Figure 6: Distribution of pupil learning in maths by performance band in treatment areas, baseline, midline, endline (%) (trends in programme areas)



Comparing the distribution of pupils by performance band over time, in Figure 6, there is a clear shift towards the top performance bands, away from the bottom bands, with the middle band similar in size to the baseline share. The most strongly significant changes over time occurred at the top and bottom of the distribution. The share of pupils with below emerging Standard 1 level skills (bottom band) fell by just under five percentage points, to reach 9% by endline. Mirroring this change, the share of pupils able to demonstrate Standard 2 level skills (top band) rose by five percentage points, to reach 9%, close to the logframe target for 2018 of 10%.

The positive change over time in maths performance is starker when considering the absolute number of pupils achieving at each level. In the 17 districts represented by the evaluation survey sample, estimates of the number of pupils in government primary schools achieving at Standard 2 level in maths increased from 6,046 pupils at baseline to 18,753 pupils at endline—more than a three-fold gain. Chapter 5 in Volume II contains the details of this analysis.

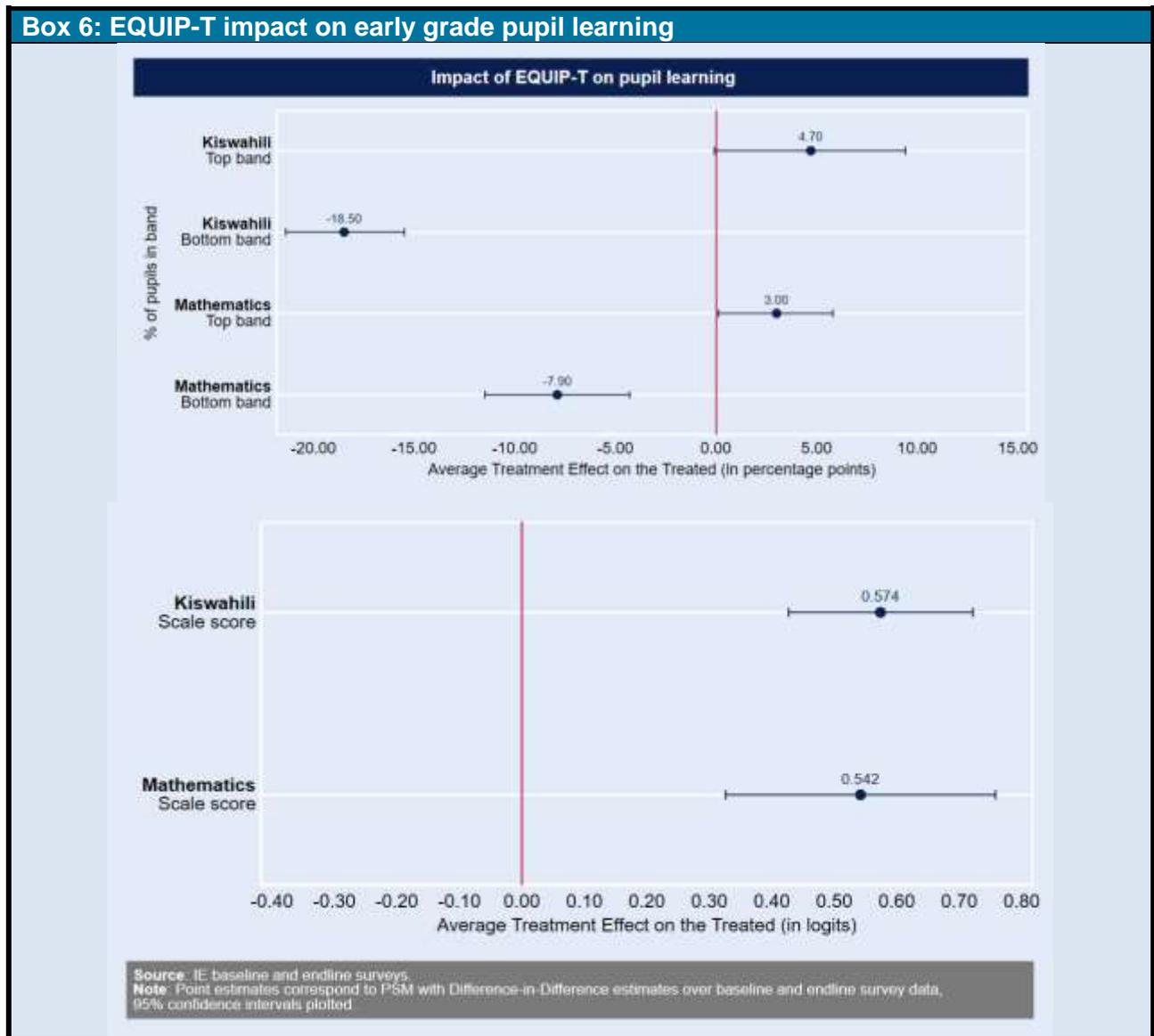
Nonetheless echoing the point made at midline, with more than 60% of Standard 3 pupils estimated to be achieving at Standard 1 level or lower, it is clear that the majority of pupils will need further support to catch up with curriculum expectations in maths, and to mitigate against the risk of falling further behind.

The disbursement-linked indicator of early grade maths achievement in the Government's EPforR program is the average percentage of correct answers on level 2 addition and subtraction

questions obtained by pupils at the end of Standard 2.⁴⁰ Standard 3 pupils in programme schools (four months into the school year) scored 25% on average at baseline on these types of questions. This increased to an average score of 30% at midline, but then did not change significantly in the two years between midline and endline (see Volume II, Chapter 5). Although the gain in numeracy skills is positive overall, it is far short of the Government’s benchmark of 80% achievement in this indicator of early grade maths skills.⁴¹

3.3.2 What is the impact of EQUIP-T on early grade pupil learning?

After four years of EQUIP-T programme interventions, the impact modelling finds strong evidence that the programme has had a positive impact on pupil learning in Kiswahili and in maths. The programme has helped pupils with both weaker and stronger skills in both subjects, and, as a result, has also had a positive impact on average scale scores. The details are in Box 6.



⁴⁰ Level 2 addition and subtraction questions are intended to test skills in the Standard 2 curriculum.

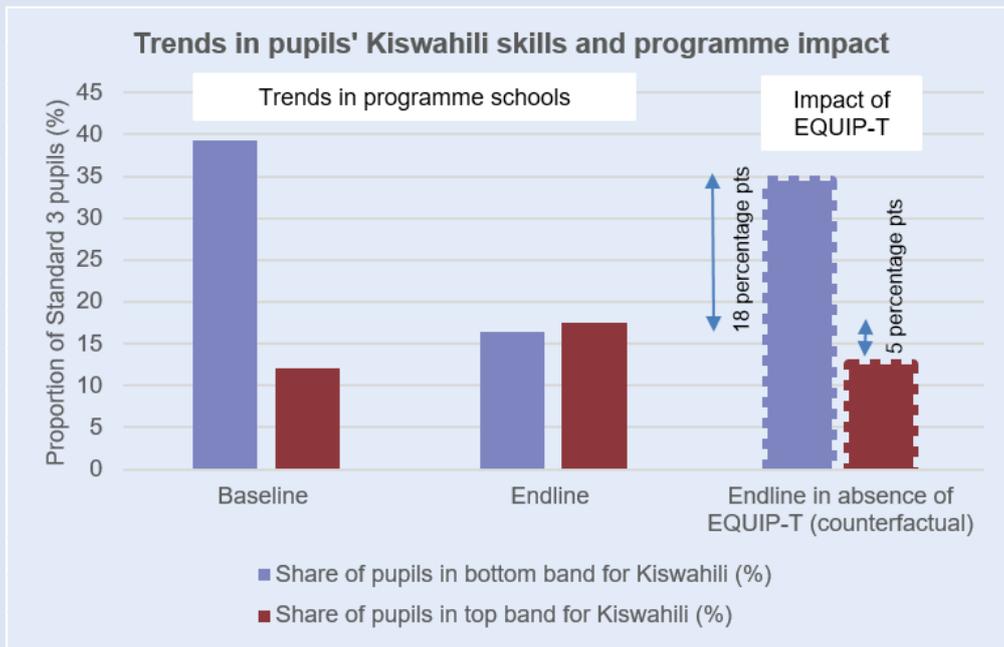
⁴¹ MOEST and PO-RALG 2018b, p.153.

The figure above shows the ATT (a measure of programme impact) on the proportion of Standard 3 pupils in the top and bottom performance bands for Kiswahili and maths (in percentage points) and, separately, the ATT detected on Standard 3 pupils' scale score for Kiswahili and maths (in logits). It compares changes in EQUIP-T schools with changes in control schools between baseline and endline.

Positive impact on pupils' performance in Kiswahili

There is strong evidence that EQUIP-T has increased the proportion of pupils in the top performance band and reduced the proportion of pupils in the bottom performance band for Kiswahili in programme schools. These results remain strong and significant across an array of estimation models and robustness checks. Pupils in EQUIP-T schools are found to be about five percentage points more likely to be in the top performance band and 18 percentage points less likely to be in the bottom performance band compared to the counterfactual situation. Similar to the impact assessments at midline, the counterfactual situation is not simply the absence of EQUIP-T treatment, but includes other, mainly training, interventions in control schools delivered by other programmes. Hence the treatment effect measures the additional effect of EQUIP-T over and above any other existing interventions taking place in control schools. For simplicity, the counterfactual situation is described as 'the absence of EQUIP-T' in the results that follow.

Given the results presented in the descriptive analysis in Figure 5 above, which indicates that 18% of pupils are in the Kiswahili top band and 16% are in the Kiswahili bottom band at endline, it can be inferred that in the absence of EQUIP-T the proportion in the top band would have been around 13%⁴², while the proportion in the bottom band would have been well over 30%. This estimated programme impact is illustrated in the chart below, alongside the overall descriptive trends in pupils' Kiswahili skills found in programme schools.



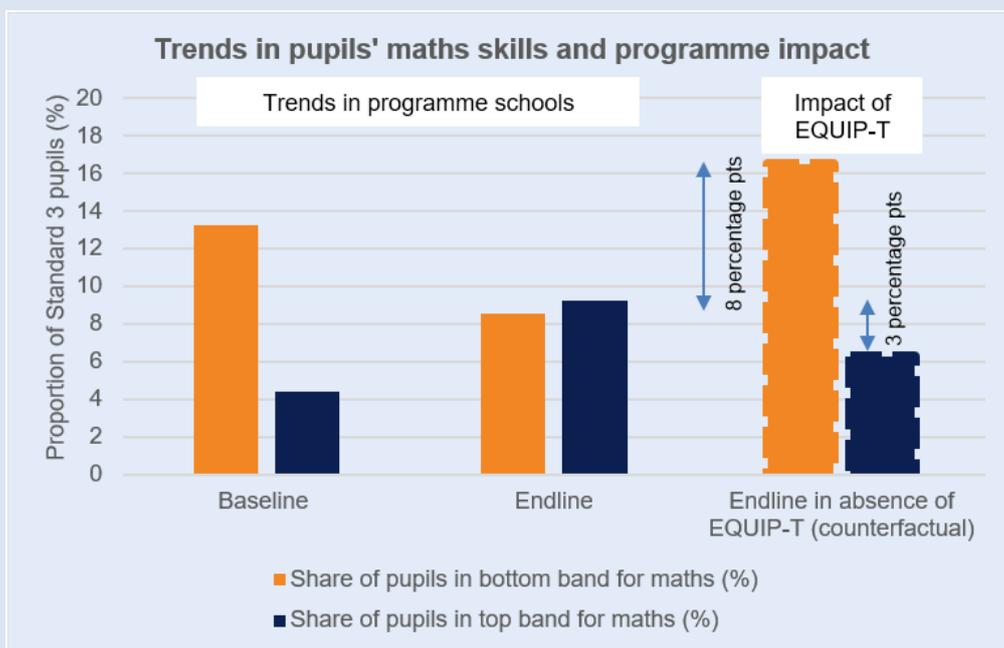
The positive impact on Kiswahili is also confirmed by the strong evidence that EQUIP-T has increased the average scale score for Kiswahili in programme schools by 0.6 logits (0.5 SD). This impact analysis at endline compares positively to the midline results, which only showed a statistically significant reduction in the proportion of pupils in the bottom performance band, but no other impact on Kiswahili skills.

⁴² This figure (13%) is calculated by subtracting 5 percentage points, which is the estimated measure of impact attributable to EQUIP-T, from 18%, which is the proportion of pupils found to be in the Kiswahili top band at endline. The same calculation is employed as part of the narrative on impact used for all other impact indicators.

Positive impact on pupils' performance in maths

There is strong evidence that EQUIP-T has had a positive impact on pupils' maths skills. The programme has increased the proportion of pupils in the top performance band and reduced the proportion of pupils in the bottom performance band for maths in programme schools. Pupils in EQUIP-T schools are found to be three percentage points more likely to be in the top performance band and eight percentage points less likely to be in the bottom performance band compared to the counterfactual situation.

Pupils' average maths score in programme schools is also higher by 0.5 logits (0.3 SD) than it would have been in the absence of the EQUIP-T intervention. On the basis of the descriptive findings in Figure 6 above, showing the distribution of pupil learning in maths, it can be inferred that in the absence of the EQUIP-T intervention the proportion of pupils in the top band would have been around 6%, whilst the proportion in the bottom band would have been around 17% (almost double the measured 9%). These results are significant across an array of estimation models and robustness checks, which gives confidence in the strength and reliability of the findings. This estimated programme impact is illustrated in the chart below, alongside the overall descriptive trends in pupils' maths skills found in programme schools.



These results are highly significant in statistical terms, as indicated by the confidence intervals not overlapping with zero in the top figure in this box, and present a consistent picture across the different dimensions through which programme impact on maths skills was measured, including maths scale score and performance bands. EQUIP-T is pushing pupils upwards in terms of maths score, which translates into pupils being lifted out of the bottom performance band and more pupils falling within the top band, with improvements visible at both ends of the numeracy outcome. These findings differ greatly from the midline analysis, which showed no clear impact on pupils' performance in maths. This seems to indicate that the programme needed the full length of the evaluation period (from baseline to endline) to achieve its impact.

Comment on impact size

A logical question which flows from this analysis of programme impact on early grade learning, is whether a 0.5 SD impact on Kiswahili scores, and a 0.3 SD impact on maths scores over four years, can be considered a low, moderate or large impact? It is very common in the education impact literature to cite impact sizes in SD units and to compare these across different impact evaluations. For example JPAL (2014) describes an effect size of 0.5 SD or more on student learning as 'very large'. Another approach is to be more specific, and to find comparators from impact evaluations of programmes that are targeting early grade literacy and numeracy in similar contexts. For example, EQUIP-T's impact results for Kiswahili are not dissimilar to the impact size of 0.3 SD found for Liberia's Teacher Training Program II on oral reading fluency or to the results from Uganda's School Health and Reading Program which found an estimated

impact on oral reading fluency of between 0.2 SD and 1.2 SD for different districts (Gove et al, 2017). But there is a question of how valid such comparisons are.

Singh (2015a) highlighted a number of flaws with using SDs to compare effect sizes across impact evaluations in education. One key problem is that a SD is a measure of dispersion, and this is not the same in different samples. This means that an intervention delivering the same absolute gain in learning would look more effective (have a higher effect size in SDs) in a context of a narrow spread in test scores than in one with a wide spread of test scores. Differences in test design, scoring, and analysis methods can also greatly affect the level of impact expressed in SD units. The message from this critique is that comparing effect sizes in SDs across different studies is not a very reliable approach. (For a more detailed discussion see Volume II Chapter 4 Section 4.6).

The impact of EQUIP-T is best understood by reflecting on the local context, on the increased share of pupils that are achieving at higher band levels as a result of EQUIP-T, and the estimated additional number of children that are reaching the required curriculum standards.

3.3.3 Why have there been gains in pupil learning overall, and what are the most likely explanations for EQUIP-T's impact?

The overall trend in learning achievement in Kiswahili and maths in the programme schools over the four years can be summed up as improving between 2014 and 2016, and then increasing only modestly (Kiswahili) or remaining stable (maths) between 2016 and 2018. Only some of the gains made over the period can be attributed to EQUIP-T, in line with the findings in Box 6, while the rest are at least partly explained by common trends experienced across the country.

At the time of the midline report, the findings from a nationally representative study (3Rs-EGRA and 3Rs-EGMA⁴³), which uses similar instruments to the tests used in the impact evaluation, found significant learning gains in Kiswahili skills, and in some (but not all) maths skills across the country over the 2014–2016 period for early grade pupils. Broadly in line with this national trend, estimates from the impact evaluation control sample also reveal significant improvement in both Kiswahili and maths skills for pupils.⁴⁴

The midline report suggested that the national trend in early grade learning improvement between 2014 and 2016 was likely to be related to the introduction of a new curriculum for Standards 1 and 2, which drastically reduced the number of subjects being taught, and markedly increased the prescribed number of instructional hours per week for Kiswahili in particular, but also maths. The new curriculum is competency-based, and it requires teachers to use a phonics approach to teaching reading, which was new to many teachers in Tanzania. The other national factor which emerged from the midline research was *hapa kazu tu* (interpreted as 'just work')—the slogan introduced by the then new government, which was credited by respondents across the midline qualitative work with influencing teachers' work ethic, and also increasing the level of oversight and monitoring of schools and teachers.

On top of the improvement in learning achievement that EQUIP-T programme schools had in common with other schools in the country over the 2014–2016 period, EQUIP-T had an additional positive impact on the achievement of the lowest performing pupils in Kiswahili (but not in maths). To explain this, the midline evidence pointed to the emphasis of early programme implementation on in-service training for teachers in Kiswahili, which had high participation rates among early grade teachers. Consistent with expectations in regard to the training, some important aspects of teachers' classroom performance in programme schools also improved significantly over the

⁴³ Research Triangle International (RTI) (2014) and RTI (2016).

⁴⁴ The control estimates are unweighted, because of the way the control group was selected, and cannot be directly compared with trends in the treatment group to infer impact (as explained in Volume II, Chapter 4, on impact estimation methods).

period, notably the use of inclusive practices. EQUIP-T has also had a positive impact on reducing teachers' absenteeism from the classroom, and thus increasing actual instructional hours.

Turning to the midline to endline period, from 2016 to 2018, results from the national 3Rs-EGRA and 3Rs-EGMA survey, conducted in late 2017, reveal significantly poorer national performance since 2016 in Kiswahili and maths skills among early grade pupils.⁴⁵ Both average Kiswahili reading speed and average score in addition and subtraction problems fell back to baseline levels. Consistent with this trend, there was a significant fall in Kiswahili and maths skills for pupils in the impact evaluation's control schools between midline and endline. The evidence suggests that EQUIP-T has helped to prevent a similar learning decline in its programme schools.

The pattern of EQUIP-T impact on early grade learning achievement over time, set out in Box 6, is consistent with the staged implementation of the teacher performance component of EQUIP-T. It is clear from the structure of the programme theory of change that this component, which principally aims to improve teaching skills in regard to the early grade curriculum competences, is the most direct link to improving pupil learning. The other components provide supporting interventions to help deliver changes at classroom level, and, to a lesser extent, to mitigate some of the demand-side barriers.

The impact of EQUIP-T on pupils' Kiswahili skills was only detectable for the lowest performing pupils at midline, while by endline there was impact on pupils with weaker and stronger skills, and on average scores. EQUIP-T's in-service training for early grade teachers on Kiswahili was largely completed between baseline and midline, but there was some continuation into the midline to endline period to complete the modules. The emphasis of the in-service model is on continuous professional development via school-based sessions and classroom practice (with peer support and feedback) in order to put theory learnt in the modules into practice. The results from the endline survey, described in detail in Chapter 4, suggest that schools are sustaining the school-based sessions over time, albeit not quite as intensively as recommended. Participation in the in-service training among early grade teachers is also reasonably high.

At midline, there was no detectable impact of EQUIP-T on early grade pupils' maths skills, yet by endline the results show impact on pupils with weaker and stronger maths skills, and on average scores. This is logical when considering the timing of the roll-out of EQUIP-T's maths in-service training for early grade teachers, which only started after midline. All nine early grade maths modules were covered by the teacher in-service training programme after midline.

While the impact of EQUIP-T on early grade learning is consistent with the roll-out of the key teacher-related interventions, the results from the endline lesson observations in treatment schools paint a mixed picture of improved and worsening teaching practices since midline. This may not be inconsistent with the modest growth (Kiswahili) or no change (maths) in learning achievement seen in treatment schools over the same period, considering the huge growth in class sizes, in particular, which makes some types of participatory teaching practices very challenging. It is notable that the significant gains in inclusive teaching practices seen between baseline and midline in programme schools were sustained at endline, despite the sharp increase in classroom overcrowding.

⁴⁵ MOEST and PO-RALG (2018b, pp. 76-79).

3.4 Trends in early grade learning gaps by gender in programme areas

Relative to boys, girls have improved their Kiswahili and maths skills at a faster rate over the last four years, and girls now significantly outperform boys in Kiswahili. At baseline there was no significant difference in average scale scores in Kiswahili by gender, but by endline there is a strongly significant gap in favour of girls. This growing performance advantage for girls in Kiswahili is also reflected in Table 4, which shows that by endline the share of girls who fall in the top performance band is eight percentage points higher than the share of boys achieving at this top level. This is mirrored at the bottom end of Kiswahili performance, where the share of girls achieving in the lowest band is eight percentage points lower than the share of boys falling in the same band. At baseline there was no significant difference in the share of pupils achieving at the top and bottom Kiswahili band levels by gender, so there has been a very marked shift over time.

Table 4: Proportion of pupils in bottom and top performance bands for Kiswahili and maths at baseline, midline, and endline by gender (%) (trends in programme areas)

	Baseline		Midline		Endline	
	Boys	Girls	Boys	Girls	Boys	Girls
Kiswahili						
Band 0 below Std. 1(%)	38.9	39.8	25.9	20.5*	20.8	12.4***
Band 2A achieving Std. 2(%)	10.9	13.2	18.0	26.7***	13.6	21.1**
<i>N</i>	(717)	(770)	(723)	(740)	(709)	(778)
Maths						
Band 0 below Std. 1(%)	10.0	16.1***	11.8	10.7	8.1	8.9
Band 2A achieving Std. 2(%)	6.4	2.7***	8.5	5.5	10.8	7.8*
<i>N</i>	(721)	(774)	(734)	(749)	(716)	(783)

Sources: Evaluation baseline, midline, and endline surveys (Standard 3 pupil tests).
Notes: (1) Asterisks indicate statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Girls started out at baseline performing significantly worse than boys in maths on average, and while this gap in average scale scores has persisted, it narrowed by midline but then did not change further by endline. In line with this overall trend, Table 4 shows that the share of girls falling into the bottom performance band at baseline was six percentage points higher than the comparative share of boys (a significant difference), but this gender disparity had been eliminated (no significant difference) by endline. This is not the case for top performers, where the gender disparity from baseline is still evident at endline. The share of boys achieving at the top performance level is 3–4 percentage points higher than the comparative share of girls at both baseline and endline, although by endline this difference is only weakly significant.

National trends in early grade learning disparities by gender between baseline and midline, based on the 3Rs-EGRA surveys (discussed above), are consistent with trends, over the same period, in the programme schools for Kiswahili, and to some extent for maths. The national results show girls gaining a significant performance advantage over boys in Kiswahili, in line with the trend in programme schools. For maths, while both the national results and the evaluation in programme schools reported boys significantly outperforming girls at baseline, the disparity had been eliminated nationally by midline, while there was still a persistent, but narrower, gap for pupils in programme areas. Common trends across the country may well explain at least some of the movement in gender gaps in early grade learning over the first two years of the programme. As

yet, there is no further nationally available data on early grade learning trends by gender covering the second two years of the programme.

Setting national trends aside, there is some evidence that EQUIP-T has contributed to the comparatively larger gains in the performance of girls compared to boys since baseline, in line with programme goals. Teachers' interactions with pupils during lessons became significantly more gender-balanced between baseline and midline, and this improvement has persisted to endline, despite the large increase in class sizes, which presents substantial challenges for effective teaching. While gender-responsive pedagogy was part of the broader in-service training for early grade teachers during the first two years of the programme, a specific gender-responsive pedagogy module was delivered after midline. A high proportion of early grade teachers report having completed this module.

3.5 Trends in early grade learning gaps by home language in programme areas

Disparities in learning achievement that favour pupils who speak Kiswahili (the language of instruction) at home, compared to those who do not, are large and strongly significant across all three survey rounds. Pupils who come from homes where Kiswahili is not the main language spoken have far lower average scale scores in both Kiswahili and maths than pupils who speak the language of instruction at home.

There has been considerable progress in narrowing the home language learning gap in Kiswahili since baseline. The gap in average scale scores for Kiswahili by home language fell by more than 50% between baseline and endline, following a consistent downward trend. In line with this finding, Table 5 shows that while the home language advantage is still significant in terms of the share of pupils falling into the lowest performance band, the gap has narrowed from 18 percentage points at baseline (44% vs 26%) to 10 percentage points by endline (18% vs 8% at endline). For the best performing pupils, contrary to the situation at baseline, there is no longer a significant difference in the share of pupils achieving in the top performance band by home language.

There has been inconsistent, and more modest, progress in narrowing the home language learning gap in maths since baseline. The gap in average scale scores for maths grew between baseline and midline, but then this gap fell to below the baseline estimate by endline. Sizeable and significant differences in the share of pupils from different language backgrounds falling into the top and bottom performance bands are still evident at endline (Table 5), which clearly illustrates the persistent disparity in maths performance.

Table 5: Proportion of pupils in bottom and top performance bands for Kiswahili and maths at baseline, midline, and endline by home language (%) (trends in programme areas)

	Baseline		Midline		Endline	
	Kiswahili	Local language	Kiswahili	Local language	Kiswahili	Local language
Kiswahili						
Band 0 below Std. 1(%)	25.8	43.5***	11.8	26.6**	8.4	18.3***
Band 2A achieving Std. 2(%)	19.1	9.9**	33.9	18.9***	21.3	16.7
<i>N</i>	(329)	(1,158)	(317)	(1,141)	(278)	(1,208)
Maths						
Band 0 below Std. 1(%)	5.5	15.6***	3.6	13.7***	2.9	9.8***
Band 2A achieving Std. 2(%)	7.5	3.5	11.7	5.6**	14.1	8.1**
<i>N</i>	(330.0)	(1,165)	(320)	(1,158)	(280)	(1,218)

Sources: Evaluation baseline, midline, and endline surveys (Standard 3 pupil tests).
Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.

EQUIP-T regions are not typical of the national situation in terms of linguistic diversity. As discussed in Section 3.2.2, 80% of Standard 3 pupils in programme districts come from homes where Kiswahili is not the main language spoken. By contrast, the national prevalence of non-Kiswahili speaking households is about 33% (Uwezo, 2011). Hence the challenge of supporting pupils whose first language is not Kiswahili is much greater in the programme regions than in many other parts of Tanzania.

EQUIP-T's main strategy for narrowing home language gaps in attainment is strengthening school readiness for pre-school-age children, partly via the SRP, and, under the extended programme, via a broader set of activities, to support expansion of Government pre-primary schooling. It is beyond the scope of this evaluation to assess the effectiveness of the SRP in supporting children who do not speak Kiswahili as a mother tongue to acquire basic Kiswahili language skills earlier than they otherwise would have done. Coverage of SRP is also a factor, as SRP centres only appear to be present in about 50% of school communities in programme areas (Box 5).

Having teachers who can speak the same language as pupils who speak other local languages at home (not Kiswahili) seems likely to be a supportive factor for this disadvantaged group. The prevalence of teachers with these linguistic skills increased significantly between midline and endline. Lesson observations, however, found that teachers were rarely observed providing extra support to pupils who do not speak Kiswahili.

3.6 Trends in learning gaps by household poverty status in programme areas

Pupils from richer backgrounds have continued to outperform pupils from poorer backgrounds since baseline, and while the gap has narrowed modestly for Kiswahili, it has grown for maths. The difference in average scale scores for Kiswahili between pupils from different economic backgrounds has narrowed slightly between baseline and endline, while the comparable difference in maths has increased over the four years. Regardless of the trend, though, the absolute disparities in average scale scores are still large.

The pattern of modest decline in the learning disparity for Kiswahili by home poverty status is also visible in Table 6. By endline, pupils from poorer homes are still significantly more likely to be achieving at the bottom performance level, and less likely to be achieving at the top performance level, than their richer peers, but the gaps between the groups have narrowed over time.

Table 6: Proportion of pupils in bottom and top performance bands for Kiswahili and maths at baseline, midline, and endline by poverty status (%) (trends in programme areas)

	Baseline		Midline		Endline	
	Poorer	Richer	Poorer	Richer	Poorer	Richer
Kiswahili						
Band 0 below Std. 1(%)	43.5	36.0**	27.2	20.7	21.7	13.1***
Band 2A achieving Std. 2(%)	8.3	14.4***	17.8	25.1***	15.0	18.8*
<i>N</i>	(477)	(957)	(531)	(925)	(579)	(904)
Maths						
Band 0 below Std. 1(%)	14.0	12.4	13.6	9.9	10.3	7.4
Band 2A achieving Std. 2(%)	4.0	4.7	5.2	8.1	6.9	10.6**
<i>N</i>	(480)	(961)	(536)	(940)	(581)	(914)
Sources: Evaluation baseline, midline, and endline surveys (Standard 3 pupil tests). Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1.						

For maths, the growth in performance disparity by economic status is also shown in Table 6 by the significant difference in the proportion of pupils from richer backgrounds who are achieving at the top skills level (11%), compared to the share from poorer backgrounds (7%), by endline. Prior to endline, there were no significant differences in the share of pupils achieving in the bottom and top bands based on poverty status.

One of the potential reasons for the gaps found in achievement by economic status is pupil absence. Data collected at endline, based on the first three months of 2018, show that pupils from poorer backgrounds are absent significantly more often than their peers from richer homes (see Chapter 7). Being absent from school means missing lessons, and thus being more susceptible to falling behind in acquiring the expected skills. This is particularly critical in maths, where skills often build on each other. Pupils living further from school are also found to be absent significantly more often than their peers living closer to school. To the extent that living more remotely is correlated with poverty this may also help to explain poverty gaps in learning. Although, overall, pupil absence rates from school have fallen slightly since baseline, they are still high—with nearly 30% of pupils absent on the day of the survey (Table 3).

Another issue related to household poverty is hunger. A large majority of pupils arrive at school each day without having eaten anything, and this situation has worsened to some extent in the last two years (see Section 3.2.1). The prevalence of school feeding programmes is still very low, at 10%, so hunger may well be having a detrimental effect on pupils' ability to concentrate and learn effectively during lessons. The other negative change in pupils' economic situation over the last two years, noted earlier, is a steep rise in the prevalence of pupils working outside the household. This may mean that pupils are more tired and find it more difficult to concentrate at school, and it may also reduce the amount of time during which additional support to education can be provided at home.

Seeking to reduce pupil absence from school and hunger are not explicit targets of the EQUIP-T programme⁴⁶, but some of the structures created under community-linked components of EQUIP-

⁴⁶ These are not indicators in the programme logframe, and they are not tracked during the annual monitoring surveys.

T, particularly PTPs and JUU clubs, have focused part of their time and resources on tackling these barriers (see Chapters 6 and 7).

Finally, it is important to highlight that pupils' home language and poverty status are correlated with each other as well as with other potential indicators of learning disadvantage such as living far from school, not eating food in the morning and working outside the household. This means that part of the poverty learning gap, for example, may be related to other background factors. While establishing the relative causes of poorer learning outcomes is beyond the scope of this evaluation, simple regression analysis of learning outcomes against a variety of background factors has been carried out. The results are reported in Chapter 7, and these show that home language is significantly and independently correlated with pupil learning – that is pupils who don't speak Kiswahili as their native language are more likely to have lower performance in both maths and Kiswahili skills. Poverty status, however, was only found to have a negative but weakly significant association with the proportion of pupils in the bottom performance band for Kiswahili but not with other measures of learning achievement once other background characteristics were controlled for.

3.7 Summary of findings on pupil learning

3.7.1 Impact results

After four years of EQUIP-T programme interventions there is strong evidence that the programme has had a positive impact on pupil learning in Kiswahili and in maths. The programme has helped pupils with both weaker and stronger skills in both subjects, and, as a result, has also had a positive impact on average scores.

The pattern of EQUIP-T impact on early grade learning achievement over time is consistent with the staged implementation of the teacher performance component of EQUIP-T—the set of interventions that are most closely related to pupil learning. The EQUIP-T teacher in-service training has high participation rates among early grade teachers, and the school-level component has been implemented broadly as intended. Teachers' pedagogical practices have become more inclusive over time, but there is mixed evidence of change in other types of positive teaching practices. At the same time, the teaching environment has become much more challenging, as class sizes have increased dramatically over time.

3.7.2 Trends in EQUIP-T programme areas

Relative to boys, girls have improved their Kiswahili and maths skills at a faster rate over the last four years, consistent with the overall goal of the EQUIP-T programme. Girls now significantly outperform boys in Kiswahili, and have narrowed the gender gap in maths. The comparatively stronger learning gains by girls than boys over the period is likely to be at least partly related to more inclusive teaching strategies, and greater involvement of girls in classroom interactions—strategies that are in line with EQUIP-T's teacher in-service training, which covered gender-responsive pedagogy.

Disparities in learning achievement that favour pupils who speak Kiswahili at home, compared to those who do not, are large and strongly significant across all three survey rounds. Although the gaps in average scores are persistently large, there has been progress in narrowing the gap in Kiswahili – but less so in maths – since baseline. The share of pupils who have a teacher that can

speak the same first language as them has increased, but there is little evidence that this has translated into more supportive classroom practices.

Pupils from richer backgrounds have continued to outperform pupils from poorer backgrounds since baseline, and while the gap has narrowed modestly for Kiswahili, it has grown for maths. Explanations for why children from poorer backgrounds struggle more with learning include absence from school, and being hungry during school, consistent with patterns found in many other countries. Some of the structures created under community-linked components of EQUIP-T, particularly PTPs and JUU clubs, have focused part of their time and resources on tackling these barriers.

4 Teacher performance

The aim of EQUIP-T component 1A is to improve teacher performance. This chapter first describes the results chain for teacher performance from the programme theory of change, implementation progress between baseline and endline, and changes expected as a result of the programme activities. The quantitative findings on this component from the evaluation's baseline, midline, and endline surveys are then presented by theme and level of the theory of change. The chapter ends by summarising the key findings and also gives an overview of to what extent inputs have succeeded in achieving the intended outputs and intermediate outcomes.

Box 7 describes some of the key concepts for the EQUIP-T teacher component.

Box 7: Description of key concepts for EQUIP-T teacher component

Communities of learning (COL): Groups of peers (in this case teachers, but the term also applies to head teachers, WEOs, and other education professionals) who support each other in continuous professional development. EQUIP-T introduced this term into its continuous professional development model after midline, although some relevant activities were already happening pre-midline. For teachers, COL activities include: attending ward cluster reflection meetings – that is, meetings with teachers from other schools in the ward on a quarterly basis; and SPMs, also known as weekly meetings – that is, teacher-led meetings at the school. During these meetings, teachers are expected to reflect on in-service training and classroom practice, discuss issues and share ideas related to school improvement and pupils' performance, and receive peer support and mentoring.

In-service training coordinator (INCO): A staff member selected by the head teacher to coordinate all EQUIP-T in-service training activities and to facilitate school-based training sessions.

Early grade teachers: EQUIP-T uses this term to refer to teachers who teach Standards 1–3. The target group for most of EQUIP-T's teacher in-service training has been Standard 1 and 2 teachers but it has been open to Standard 3 teachers too.

3Rs curriculum: The Standards 1 and 2 curriculum that was revised in 2015, and focuses on reading, writing, and arithmetic (see Chapter 1, Section 1.4.1).

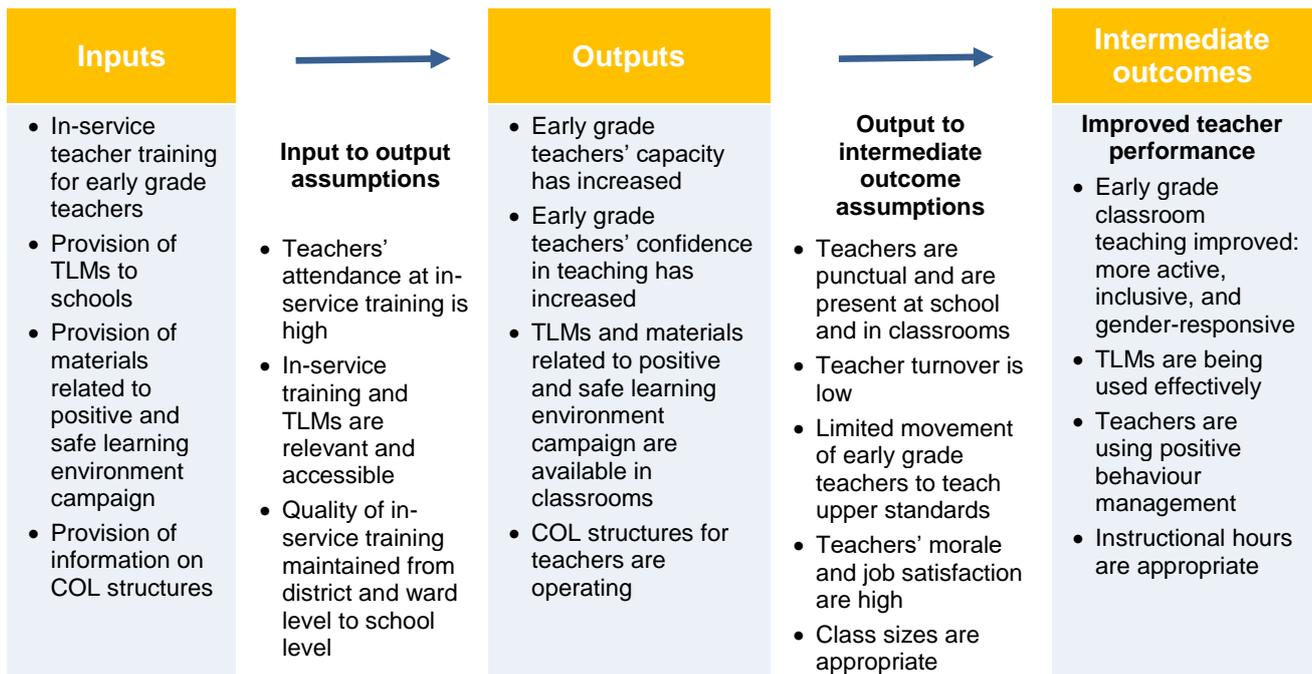
4.1 Teacher performance: theory of change, implementation, and expectations of change

4.1.1 Theory of change

Teacher performance is central to the overall EQUIP-T theory of change because the expected results from this component directly affect pupils' learning experiences. Figure 7 shows the inputs, outputs, intermediate outcomes, and assumptions in the main results chain for EQUIP-T component 1A, which are covered by the quantitative research.

Note that evidence on the receipt of materials related to positive and safe learning environments and their availability in classrooms, as well as teachers' use of positive behaviour management, are discussed in Chapter 7.

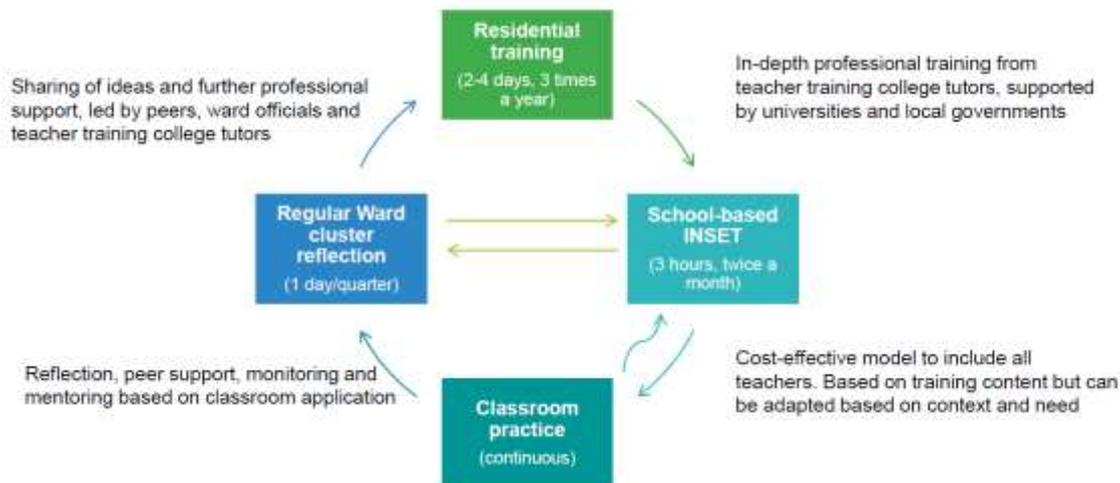
Figure 7: Main results chain for EQUIP-T component 1A teacher performance⁴⁷



4.1.2 Implementation since baseline

The largest intervention of the programme under component 1A has been the provision of in-service training for early grade teachers. EQUIP-T uses a continuous professional development model that at its core is a school-based in-service training programme.

Figure 8: EQUIP-T's current model of in-service training for teachers



Source: EQUIP-T MA, presentation at UKFIET conference, Oxford, September 2017.

Figure 8 provides a schematic representation of EQUIP-T's current training model for teachers, which has evolved over time. At the beginning of implementation, the training was based on a cascade model, whereby a small number of early grade teachers from each school, as well as the INCO, received the residential training at the district level and then cascaded the training to all other teachers through school-based training sessions. The current model, on the other hand, is

⁴⁷ See Annex C for the full endline evaluation matrix, which includes additional input to output assumptions for component 1A that are not covered by the quantitative research.

expected to provide an opportunity for all early grade teachers to receive at least some residential training, before continuing learning at the school and ward level. The model has also evolved into a COL approach that relies less on centrally developed modules and materials, and focuses more on peer learning and on teacher-led identification of issues and demand for discussions on how to teach different competencies. The current model is expected to present an affordable and sustainable model of professional development (EQUIP-T MA, 2017).

EQUIP-T has rolled out in-service training and delivery of supporting TLMs in stages since baseline. Box 8 describes these inputs, and Annex B.4 provides additional details on the EQUIP-T in-service training model and roll-out. In broad terms, the order of the in-service training up until mid-2018 was: 3Rs curriculum training and the first three sets of Kiswahili literacy training in 2015 and 2016, followed by the first set of numeracy training in 2016, then gender-responsive pedagogy training and the second set of the numeracy training in 2017. There has been regional and district variation in the roll-out of the different training sets and in the exact mode of organisation. As a result, the information on the participants and number of days for each training set, as described in Annex B.4, might differ by districts.

Box 8: Staged implementation of in-service training and delivery of TLMs

Between baseline and midline EQUIP-T provided the following four sets of teacher training and TLMs:

- **Early grade Kiswahili literacy in-service training (sets 1, 2, and 3)** targeted at all Standards 1 and 2 teachers and some Standard 3 teachers, INCOs, head teachers and WEOs. Set 1 covered modules 1–4 of the programme on literacy, while set 2 covered modules 5-8, and set 3 partially covered modules 9-13;
- **3Rs curriculum in-service training (set 4)** targeted at all Standard 1 and 2 teachers; and
- **Early grade TLMs**, including supplementary readers, 'big books', teacher 'read-aloud' books, and literacy teaching aid toolkits.⁴⁸

Since midline, EQUIP-T has provided another four sets of training and additional TLMs (see below). The concept of COL was also introduced, and schools are expected to hold **weekly SPMMs** and to send teachers to **quarterly ward cluster reflection meetings** (see Annex B.3 Table 21). The training and TLMs include:

- **continued early grade Kiswahili literacy in-service training (set 5)**, which completed modules 9–13;
- **early grade numeracy in-service training (sets 6 and 7)**, targeted at all teachers of Standards 1 and 2, and some other maths teachers, INCOs, head teachers and WEOs. Set 6 covered modules 1–4 of the programme on numeracy, while set 7 covered modules 5–9⁴⁹;
- **gender-responsive pedagogy in-service training (set 8)**, targeted at selected teachers (not necessarily from early grades), INCOs, head teachers, and WEOs;
- **early grade numeracy teaching aid toolkits**;⁵⁰ and
- **posters** on positive classroom behaviour management and a positive and safe learning environment.

For more details on the implementation of the overall teacher performance component since baseline, see Annex B.3.

4.1.3 Expected changes

The changes expected as a result of the programme by midline, according to EQUIP staff,⁵¹ were that: teachers would be more confident in their teaching and use more inclusive teaching practices; pupils would be more actively involved during lessons; and the use of home-made teaching aids and TLMs in classrooms would increase.

EQUIP-T staff's expectations were that by endline early grade teachers will be more confident and competent in teaching 3Rs; their lessons will be more interactive; and sufficient teaching aids will be available and being used frequently by teachers; while pupils will be enjoying their lessons more; feel more motivated; and attend school more often. In addition, the school-based EQUIP-T in-service training is expected to have spread to upper grade teachers.

⁴⁸ Supplementary readers means a set of reading books for pupils which have been organised into reading levels, so that pupils work gradually up the levels as their skills improve. 'Big books' and 'read-aloud' books have the same purpose – the teacher reads them out loud to the class or pupils share them in groups for peer-to-peer learning. The toolkits include posters and basic materials to enable teachers to make their own teaching aids.

⁴⁹ The EQUIP-T early grade numeracy training for teachers includes 13 modules overall, however, only the first nine out of the 13 were rolled out by mid-2018 due to delays in approval.

⁵⁰ The toolkits include posters and mathematical instruments, as well as basic materials to enable teachers to make their own teaching aids.

⁵¹ In January 2016, the evaluation team interviewed the component technical leaders and the programme's senior management team to elicit their views. In January 2018, the evaluation team held a workshop with similar participants from the EQUIP-T MA, where expected changes were discussed.

Some of these expected changes are reflected in the programme's logframe indicators for EQUIP-T component 1A. These capture early grade teachers' pedagogical practices; teachers' use of gender-responsive pedagogy in their classroom while teaching; and teachers' active engagement in in-service training groups or COL (EQUIP-T MA, 2017).⁵² For more details on logframe indicators see Annex B.6.

4.2 Findings: EQUIP-T in-service training for teachers and COL

This section presents the key findings on the delivery of the EQUIP-T in-service training for early grade teachers, as well as on teachers' attendance and perceptions of the training. Chapter 5 in Volume II presents additional findings on the school-based component of the in-service training that are not reported in this section, and on the profile of the INCOs. These supplementary findings also report in-service training attendance and module completion indicators at endline by gender and age of teachers, and find no significant differences for the different groups.

At endline, **school level** data on in-service training were collected through group interview with the INCO and other teachers at the school, in addition to **teacher-level** data from individual interviews. For the remainder of this section, findings on in-service training from both sources are presented and compared as appropriate.

4.2.1 Have teachers received EQUIP-T in-service training? (EQUIP-T input)

At endline, almost all Standards 1 and 2 teachers (98%) had attended EQUIP-T in-service training in 2016 or 2017. There is also high coverage of EQUIP-T training (90%) for teachers who teach Standard 3 but who do not teach Standards 1 and 2 over the same two years. This shows that EQUIP-T training reached most of the targeted teachers.

Among Standards 1 and 2 teachers who attended EQUIP-T training in 2016–17, 71% had attended both training away from school and school-based training sessions. Nonetheless, 17% had only attended school-based sessions, which is a significant increase (by 13 percentage points) since midline, and is at odds with EQUIP-T's evolved training model, whereby all early grade teachers are expected to receive some training away from school. The remaining 12% of teachers had only attended sessions away from school, which is a concern given that Standards 1 and 2 teachers are expected to attend all of the school-based sessions (see below).

The overwhelming majority of schools (97%) have an INCO, but turnover is very high. Only 43% of schools have had the same INCO in post since January 2015, when school-based in-service training started. Having such high turnover is particularly problematic given the INCOs' central role in supporting peer learning and mentoring of teachers, which relies on relationships of trust and respect, which tend to take time to build.

EQUIP-T in-service training away from school

All programme schools report that at least one teacher from the school had attended EQUIP-T in-service training away from school since baseline on early grade Kiswahili literacy, early grade numeracy, and gender-responsive pedagogy. On average, schools report that teachers had received the following training sessions away from school since baseline: nine days of training on early grade Kiswahili literacy in 2014 and 2015, and three days in 2016; four

⁵² Note that while this evaluation measures teachers' demonstration of positive and gender-responsive pedagogical practices, as well as their engagement with school-based in-service training and COL, these indicators are not defined in the same way as the logframe indicators.

days on early grade numeracy in 2016 and five days in 2017; and three days on gender-responsive pedagogy in 2017. Each session attended on Kiswahili literacy covered, on average, five to six modules; while sessions attended on numeracy covered about five modules per training session.⁵³ These findings are broadly in line with EQUIP-T's reporting of implementation since baseline (see Annex B.4).⁵⁴

School-based EQUIP-T in-service training

There is wide variation in the implementation of the school-based in-service training across programme schools. Schools are expected to hold training sessions for three hours twice a month, while school is in session (Figure 8). Accounting for term breaks and holidays, this would translate into about 18 sessions per year. In 2017, 45% of schools held 15 or more school-based training sessions, while 29% held from five to nine sessions, and 19% held zero to four sessions (Figure 9).⁵⁵ These sessions last on average, two hours, which is less than the stipulated three. Therefore, although a large share of schools hold school-based training as intended, the majority do not, which is likely to undermine EQUIP-T's impact on teacher performance.

Schools' implementation of the school-based in-service training has been improving since 2015. The proportion of schools that hold zero to four sessions per year has significantly dropped, from 38% in 2015 to 19% in 2017; while the proportion holding 15 or more sessions has significantly increased, from 30% in 2015 to 45% in 2017.

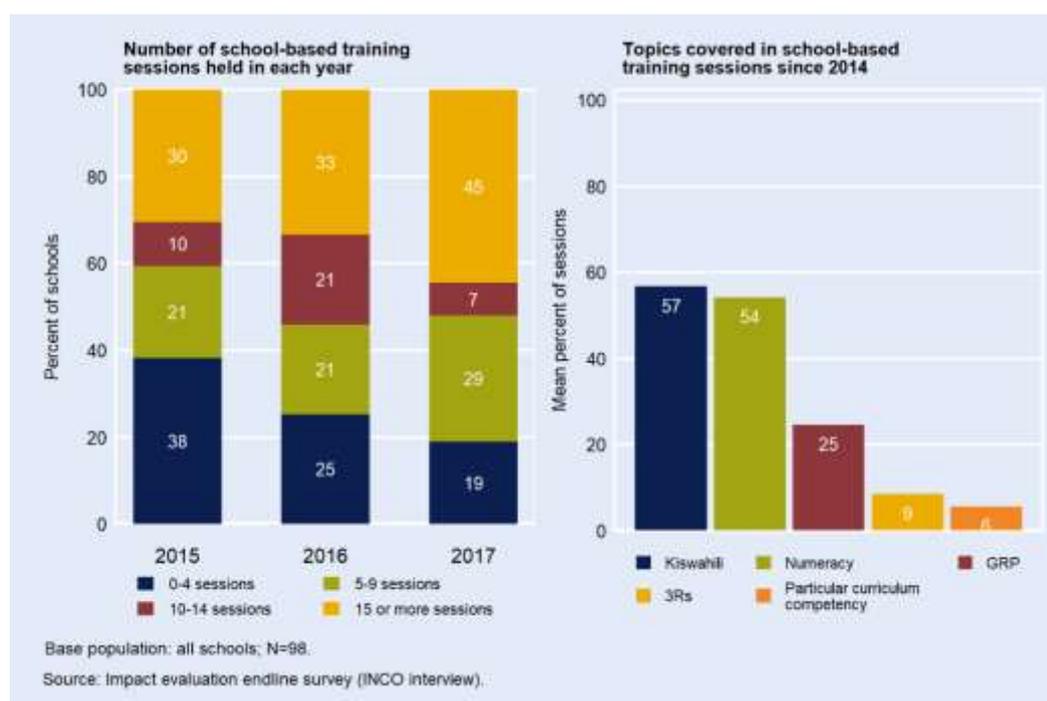
The topics covered in the school-based in-service training mostly relate to early grade Kiswahili literacy and numeracy, and are in line with the training delivered away from school. On average, 57% of all school-based sessions that took place since baseline covered Kiswahili literacy, 54% covered numeracy, and 25% covered gender-responsive pedagogy (Figure 9). This implies that teachers are covering a combination of topics per school-based session. Only 6% of all sessions since 2014 have been on particular curriculum competencies where the training takes less of a modular nature and instead teachers discuss specific competencies/skills from the curriculum and related teaching techniques. This is at odds with intended direction of EQUIP-T's training model towards teacher-generated learning and less on centrally developed modules.

⁵³ Note that these could also include modules that were covered as a refresher from the previous sessions.

⁵⁴ Some teachers confuse the training on the 3Rs curriculum with training on early grade Kiswahili and numeracy, given that both cover Kiswahili and numeracy topics from the new syllabi for Standards 1 and 2. Hence, the number of days reported by schools should not be expected to perfectly match with those reported by the programme.

⁵⁵ Note that schools do not necessarily hold two school-based training sessions per month. It could be that there are certain periods in which they hold several sessions per month and other periods in which they hold one or no sessions per month.

Figure 9: EQUIP-T school-based in-service training implementation since baseline (trends in programme areas)



Completion of the EQUIP-T in-service training modules

A large share of Standards 1 and 2 teachers report not completing all the EQUIP-T in-service training modules delivered by the programme so far. Among Standards 1 and 2 teachers, only 46% have completed all 13 Kiswahili modules as part of their training away from school and in school, 58% have completed all nine numeracy modules, and 72% have completed the gender-responsive pedagogy module. This is despite the fact that all Standards 1 and 2 teachers should have completed all these modules by the end of 2017. The reasons for many teachers failing to do so are not clear, but cannot be readily explained by teacher turnover. The majority of new early grade teachers who joined the school since baseline came from other EQUIP-T districts, and should therefore have received the training in their previous schools if they were early grade teachers there. Regardless of the cause, this will likely weaken any EQUIP-T impact on teacher performance.

At the same time, more than 85% of schools report that they have completed all the Kiswahili literacy, numeracy, and gender-responsive pedagogy modules in their school-based sessions. This large discrepancy between the reported school- and teacher-level training module completion rates suggest that some Standards 1 and 2 teachers are not attending all of the school-based sessions held by the school. This is discussed in the next two sub-sections.

Training spill-over to upper grade teachers

There is large spill-over of EQUIP-T training, with a sizeable proportion of upper grade teachers attending training away from school. On average within a school, 41% of teachers who have attended EQUIP-T training away from school since baseline do not teach Standards 1 to 3. This partly reflects the attendance of the INCO, who in 42% of schools teaches upper grade subjects only and is expected to attend all of the EQUIP-T residential training sessions. It may also be due to the attendance of the gender focal teacher for the gender-responsive pedagogy training, who can be an upper grade teacher. There could also have been some cases of mis-targeting

where schools chose other teachers to attend the training rather than the targeted Standard 1 and 2 teachers; however, the evaluation does not collect evidence to support or refute this possibility.

There is also spill-over of school-based EQUIP-T training to upper grade teachers. When it comes to the last school-based in-service training session that a school held, on average 39% of teachers attending that session were not teaching Standards 1 to 3. This is in line with EQUIP-T's expectations that by endline, school-based training would have spread to upper grade teachers.

4.2.2 Teacher attendance at EQUIP-T in-service training (EQUIP-T input to output assumption)

Many early grade teachers are not regularly attending the EQUIP-T school-based in-service training. Of all Standards 1 and 2 teachers at endline, 12% have not attended any of the school-based training sessions in 2016 and 2017. Furthermore, on average, Standards 1 and 2 teachers report attending 12 days of school-based training during 2016–17 – that is, six days per year. However, schools report holding on average 15 days of school-based training sessions in 2016 and 16 days in 2017.

This is supported by teachers' self-reporting of attendance. Among Standards 1 and 2 teachers who attended any school-based training in 2016–17, only 52% report attending all of the sessions held by the school. This is a significant decline since midline, when 66% of teachers reported attending all of the school-based training sessions.⁵⁶ This poses a serious concern as a sizeable share of the target group of teachers are not benefiting from the full in-service training model. Irregular participation in the school-based component may well reduce programme impact on strengthening teacher skills and classroom practices.

4.2.3 Relevance and accessibility of EQUIP-T in-service training (EQUIP-T input to output assumption)

Nearly all teachers of Standards 1 to 3 (99%) found the EQUIP-T in-service training in 2016–17 useful. The most important gains reported by teachers are: general teaching skills (63%); inclusive teaching skills (53%); gender-responsive teaching skills (48%); confidence in teaching (45%); and lesson planning skills (41%). A significantly larger share of teachers at endline report gaining confidence in teaching than at midline (45% compared to 30%), and similarly for gaining lesson planning skills (41% compared to 26%) and gaining a support network (13% compared to 3%).

On the other hand, more teachers at endline are facing difficulties with the EQUIP-T training than at midline – 77% compared to 56%. At school level, the main challenges reported with the EQUIP-T training away from school are: limited training time (76%), insufficient allowance⁵⁷ (65%), not enough training material (40%), and problems with transport to the training venue (25%).

At school level, the main challenges reported with school-based training are no allowance (78%) and participants not being motivated (59%), followed by limited training time (50%), not enough training materials (42%), and sessions taking place at an inconvenient time (40%).⁵⁸ The

⁵⁶ This is also supported by the school-level data, according to which 39% of Standards 1 to 3 teachers did not attend the last school-based training session that the school held.

⁵⁷ The endline survey found that 11% of Standards 1 to 3 teachers had paid out-of-pocket money to attend the EQUIP-T training away from school in 2016–17.

⁵⁸ Challenges with EQUIP-T training reported by Standards 1 to 3 teachers in the teacher interviews are consistent with the challenges reported at the school level. Volume II, Chapter 5, presents the challenges reported by teachers.

high shares of schools reporting no allowance and low motivation of participants as challenges with the school-based training is concerning as the sustainability of the EQUIP-T in-service training model relies heavily on teachers' motivation to regularly attend and learn from the training. The lack of training materials is another of the main concerns for teachers, although EQUIP-T is evolving the in-service training model away from centrally developed materials.

The reported challenge of sessions taking place at inconvenient times could in part explain the low attendance of Standards 1 and 2 teachers at the school-based training. The majority of schools (86%) hold the school-based training sessions on school days after teaching hours (as intended in the model), while 11% hold the sessions during teaching hours. The latter might be contributing to teachers' absence from their classrooms, and thus could be an unintended negative consequence of the programme. Schools decide when to hold school-based training sessions, yet just over one-fifth of schools (22%) said that holding training when schools were closed would improve the delivery of EQUIP-T in-service training.

4.2.4 Are teacher COL structures operating? (EQUIP-T output)

Not all schools are participating in ward cluster reflection meetings, and among those that are, attendance is much lower than expected. A sizeable minority of schools (23%) did not have any teacher attending a ward cluster reflection meeting in 2016 and 2017. Among schools that did participate in these meetings, on average, teachers attended only three days of ward cluster reflection meetings in 2016–17. This is low, given that these meetings are expected to take place on a quarterly basis. The meetings mostly took place on school days during teaching hours (76% of schools) and typically lasted about six hours per day. The main topic of discussion at the last meeting was on numeracy for the majority of schools (54%), followed by a general discussion and evaluation of the in-service training modules (28%).

SPMMs are taking place in most schools but not always on a weekly basis and for their intended purpose. Of all Standards 1 to 3 teachers, 83% report attending at least one SPMM in the last 60 days, and half of these teachers report attending four or more over the same period. However, 25% of teachers attended two or fewer meetings in the last 60 days. When asked at the school level, only 34% of head teachers report that the school held four or more SPMMs in the last 60 days, clearly indicating that many schools are not holding SPMMs on a weekly basis. Moreover, the SPMMs are intended to cover discussions on classroom teaching, teacher in-service training, teacher or pupil attendance, or pupil learning. However, only 65% of schools mainly discussed any of these topics at the last SPMM.

4.3 Findings: teacher capacity and confidence

4.3.1 Has teacher capacity and confidence improved? (EQUIP-T output)

Teachers' confidence in teaching the new Standards 1 and 2 curriculum is high but there has been no change since midline. At endline, 78% of Standards 1 and 2 teachers are very confident with teaching the new 3Rs curriculum, while 22% are fairly confident, and less than 1% are not confident. At midline, these shares were similar and teachers in the midline qualitative research associated their confidence with the EQUIP-T training on the 3Rs curriculum (OPM, 2017a).⁵⁹ However, it is a concern that over 22% of teachers are still not very confident with

⁵⁹ 79% of Standards 1 and 2 teachers at midline reported that the EQUIP-T training they had received in 2014–15 included training on the 3Rs curriculum. It is worth noting that this share could possibly be higher or lower as teachers confuse training on 3Rs with training on early grade Kiswahili literacy and numeracy.

teaching the new curriculum, especially given they have had two more years of teaching experience since midline.⁶⁰

The evaluation does not measure teachers' knowledge of positive and inclusive teaching practices, instead it measures demonstration of these practices, as reported in Section 4.5. Another aspect of teachers' capacity is subject knowledge. The midline evaluation found no change in average levels of teachers' maths or Kiswahili subject knowledge since baseline, when the average score on both assessments was about 60% (OPM, 2017a).⁶¹ This was not measured again at endline, partly because EQUIP-T's in-service training does not focus on this aspect of capacity.⁶²

4.4 Findings: class size, teacher turnover, and job satisfaction (EQUIP-T output to intermediate outcome assumptions)

Before presenting findings on teaching practices in Section 4.5, it is first useful to see from the findings below that some of the conditions which support the link between delivering in-service training and better teaching practices have worsened significantly since midline.

4.4.1 Class size

Class sizes have been markedly increasing since baseline and midline, reaching critically high levels at endline (Figure 10). The average class size for Standard 1 grew significantly from 75 pupils at baseline to 104 pupils at endline, and, similarly, the average Standard 2 class size is at 109 pupils, an increase of 68% since baseline and 52% since midline. While average class size is lower for Standard 3, it is still high, at 88 pupils, with significant increases since both baseline and midline. These trends are likely driven by the Government's fees-free basic education policy, introduced in December 2015, as well as a change in the age of entry to primary school from seven to six years.

There is also considerable variation in class sizes across schools. For example, while the 10% of schools with the lowest class size for Standard 2 have 53 or fewer pupils, at the other end of the distribution, 10% of schools have at least 193 pupils, more than three times the official target class size of 60 pupils (MOEST, 2018).⁶³ Overcrowded classrooms, while beyond the influence of EQUIP-T, can severely undermine teachers' use of effective teaching practices, including those learnt during the in-service training.

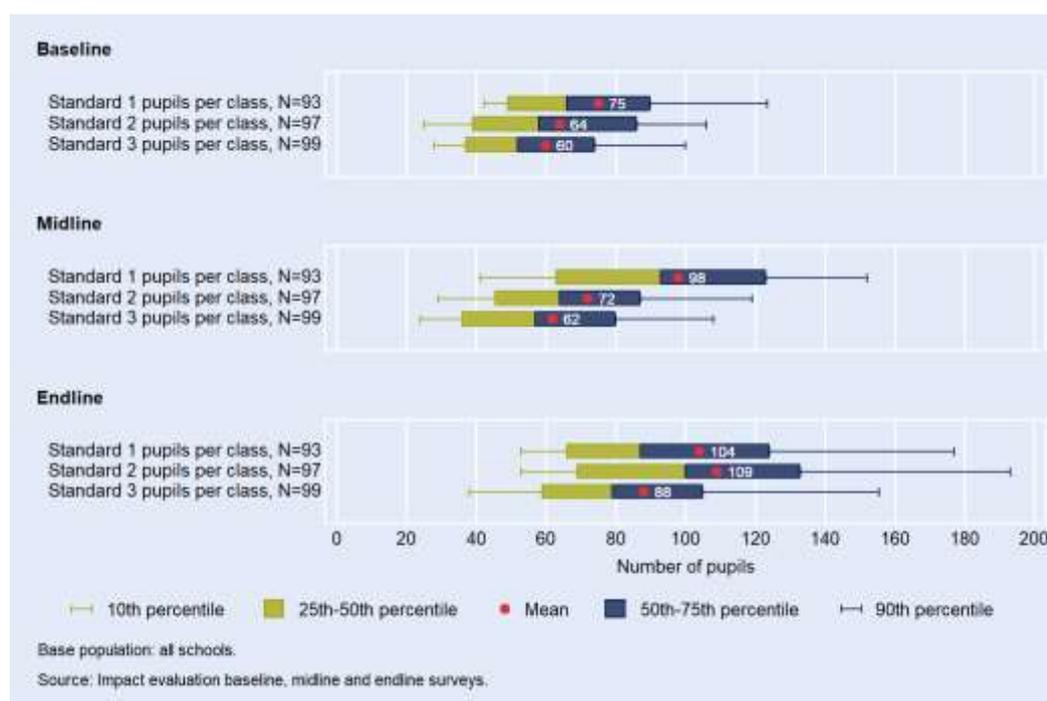
⁶⁰ The data show that teachers who are fairly but not very confident with teaching the curriculum are not necessarily those who have recently joined the teaching profession but also include teachers with long teaching experience who will have received the training on 3Rs in 2015.

⁶¹ Teachers' subject knowledge was assessed using a teacher development needs assessment instrument which contains questions linked to the primary school curriculum and which is in the form of a mock pupil test that teachers mark, indirectly providing information on their subject knowledge. For further details on the instrument development, refer to OPM (2015b), Section 3.3.

⁶² See Volume II, Chapter 3 for further explanation of the reasons why subject knowledge was not assessed again at endline.

⁶³ MOEVT (2009a) states that 40 pupils per class is the benchmark for primary classes; however, the latest ESDP stipulates an interim target of 60 pupils per class (MOEST, 2018).

Figure 10: Class sizes by standard: baseline, midline, and endline (trends in programme areas)



Schools have not been able to recruit the increased number of teachers needed to cope with the rise in enrolment. While the official acceptable range of the pupil–teacher ratio is between 35 and 53 (MOEST and President’s Office Regional Administration and Local Government (PO-RALG), 2018), at endline there are on average 66 pupils for every teacher, and this is a significant increase from 54 at baseline and 51 at midline.⁶⁴

Despite the growth in class sizes, the vast majority of pupils have a useable desk at endline and this share has increased significantly since baseline and midline.⁶⁵ On average, 92% of pupils in the observed Standard 2 lessons had a useable desk, which is a notable increase of 21 percentage points since baseline and 15 percentage points since midline. This is most likely due to the Government policy that was announced in 2016 and that required that each pupil have a desk space.⁶⁶ While this initiative has led to substantial improvements in desk availability in schools, there are still 8% of pupils who do not have a desk space and this share would be higher if all enrolled pupils attended their lessons on a given day (see Section 3.2.2 in Chapter 2 for estimates of pupil absenteeism).

4.4.2 Level of teacher turnover

Teacher turnover in this report is defined as the rate in which teachers leave their current posts that are defined by both their role and place of work. Therefore, teachers who remain in the same role but change their place of work, such as teachers who transfer to another school, are included

⁶⁴ This is for all pupils and teachers of Standards 1 to 7.

⁶⁵ The term ‘useable desk’ means a space at a desk where a pupil can sit and write without being cramped. Often desks are designed for more than one pupil.

⁶⁶ In 2016, the Government of Tanzania gave an order to regional and district commissioners to ensure that all schools have enough desks for their pupils. <http://www.mwananchi.co.tz/Maagizo-sita-ya-Magufuli/1596774-3119140-9ho24j/index.html>

in the turnover rate. Similarly, teachers who remain in the same school but change roles, such as get promoted to head teacher, are also included in the turnover rate.

Teacher turnover is still very high. Of all Standards 1 to 3 teachers at midline, 25% are no longer at the same school by endline, not significantly different to the turnover rate of 28% between baseline and midline. Looking at turnover over a four-year period, almost half of all Standards 1 to 3 teachers (47%) at baseline are no longer working at the same school at endline.

In addition to the large shares of teachers leaving their schools, there is also movement of teachers between the standards being taught. Of all Standards 1 to 3 teachers at baseline, 12% are still at the same school at endline but are no longer teaching early grade standards. Combining turnover with movement to upper standards leads to only 41% of Standards 1 to 3 teachers at baseline who are still teaching these standards in the same schools at endline.

While 56% of teachers who left their school since midline are still teaching or working in education,⁶⁷ 44% of them have left the teaching profession altogether. The main reasons for the teacher turnover between midline and endline are: transferring to another school (48%), retiring (23%)⁶⁸, disciplinary issues (9%)⁶⁹, seeking further studies (9%), and quitting their job (7%). For teachers leaving the profession, any skills and knowledge gained from the EQUIP-T in-service training will no longer be of benefit to pupils, and thus dilutes the possible impact of the programme on learning.

All these findings taken together point towards severe constraints on teachers and pupils benefiting in full from the EQUIP-T in-service teacher training. While the overwhelming majority of Standards 1 to 3 teachers who joined their schools between midline and endline come from other EQUIP-T districts and regions, the change in the standards that teachers teach over time, and the high turnover of teachers on a regular basis, is disruptive to the effective delivery of the in-service training. This is particularly the case for the EQUIP-T in-service training model, which relies to a large extent on school-based peer learning that is continuous and sustainable. For more details on teacher turnover, see Volume II Chapter 5 Section 5.4.

4.4.3 Teacher job satisfaction and motivation

The quantitative component of the evaluation does not directly measure teachers' motivation or morale. While absence from school and classrooms, tardiness (Section 4.7.1), and reduced efforts to prepare for lessons (Section 4.5.2) may be linked to low teacher motivation, it may also reflect practical constraints. Low levels of teacher motivation have the potential to affect teachers' engagement with in-service training and to undermine the link between this and improved teaching practices. EQUIP-T aims to improve teacher morale and motivation, implicitly through its continuous professional development model, and via its support to the national teacher continuous professional development framework (EQUIP-T MA, 2015).⁷⁰

⁶⁷ This may be overestimated as it assumes that all teachers who left to seek further studies will return to the teaching profession, which might not be the case.

⁶⁸ Of all Standards 1–3 teachers, 5% will reach the official retirement age of 60 years in the next year.

⁶⁹ This is most likely associated with the mass dismissal of teachers that took place nationwide in 2017 due to the discovery of fake education certificates by some teachers.

⁷⁰ EQUIP-T's original design had an explicit teacher motivation and morale intervention called a toolkit, but this was never implemented.

There has been a slight improvement in teachers' reported job satisfaction since baseline but no change in their perceived appreciation by the community and head teacher.⁷¹ At endline, Standards 1 to 3 teachers self-report a mean score of 8.1 out of 10 for job satisfaction, a significant increase of 5% since baseline but not a significant change since midline. They report a mean score of 8.4 for head teacher appreciation of their role and a mean score of 6.4 for community appreciation of their role, both unchanged since baseline or midline. This shows that reported job satisfaction is still relatively high on average⁷² and teachers still tend to feel less valued by the community than by their head teacher.

However, when asked how their level of job satisfaction compares to two years ago, there has been a negative change since midline. While over half (55%) of Standards 1 to 3 teachers at midline reported that their job satisfaction was higher than two years ago, only 36% of teachers at endline report this being the case, a significant and large drop.

Many factors can negatively affect teachers' motivation and morale. The huge growth in class sizes since midline is likely to have substantially increased teachers' workload in preparing for lessons and marking, and to have made teaching itself more challenging. Tackling these conditions may have made some teachers feel overwhelmed and discouraged. The baseline and midline qualitative research highlighted teachers' views that the shortage of any and good quality housing for them was a major problem (OPM, 2015a; OPM, 2017a). While the endline survey has not explored this further, it is unlikely that the situation has substantially improved as the proportion of schools without a single teacher house⁷³ and the average time to school for teachers⁷⁴ remain largely unchanged since midline. Being paid non-salary allowances late is a problem for a majority of teachers. At endline, 70% of Standards 1 to 3 teachers have outstanding non-salary claims – for example for leave, transfer, or sickness.⁷⁵ While most of these supporting conditions are not within the control of EQUIP-T, they can greatly influence programme success.

4.5 Findings: classroom teaching

This section presents evidence on the key intermediate outcome in the theory of change of component 1A: improved classroom teaching. It examines whether classrooms are more active in terms of teaching and learning, and whether they are more inclusive and gender-responsive; and whether the use of positive teaching practices and continuous assessment have improved over time. Supplementary findings on the use of teaching practices by gender and age of the teacher are in Chapter 5 in Volume II. At the endline, there are very few significant differences in the classroom practices observed for the different groups.

Enumerators carried out two types of observations in Standard 2 Kiswahili and maths lessons: first, a mapping of teacher–pupil interactions by gender and by classroom space; and, second, an observation of teachers' use of a set of selected teaching behaviours (Annex E).⁷⁶ These form the

⁷¹ On the day of the survey, teachers were asked how satisfied they were with their jobs and how appreciated they felt for their work by the community and head teacher, by placing themselves on a 10-point scale, where 1 was 'completely unsatisfied' and 10 was 'completely satisfied'.

⁷² It is important to note, however, that this mean score may be overestimated as teachers may be reluctant to report truthfully about their job satisfaction out of fear that this might be reported to their head teachers.

⁷³ 11% of schools at endline, compared to 13% at midline, with the difference not being significant at the 10% level.

⁷⁴ At endline, it takes teachers 18 minutes on average to get to the school, compared to 15 minutes at midline, with the change not significant at the 10% level.

⁷⁵ This information was not collected at baseline or midline and therefore it is not possible to examine the trend.

⁷⁶ The lesson observation instrument was adapted from tools used to evaluate a school-based in-service teacher training programme in Tanzania (Hardman and Dachi, 2012).

basis of the evidence on teachers' use of inclusive, gender-responsive, and positive teaching practices in the classroom.

4.5.1 Has the use of inclusive and gender-responsive teaching practices in the classroom increased? (EQUIP-T intermediate outcome)

The gender balance of teachers' interactions with pupils in the classroom has improved significantly since baseline but is unchanged since midline. On average, 68% of teachers' interactions with pupils are gender-balanced – that is, teachers engage proportionally with boys and girls present in the classroom (Table 7).⁷⁷ This is a significant increase of 15 percentage points from baseline, which is mainly due to the significant decline in the proportion of lessons where teachers interact more with boys. Despite this improvement, girls are still more disadvantaged than boys and there has been no significant change in the gender balance of teacher interactions since midline. This indicates that the improvement happened between baseline and midline and was then sustained at endline, despite the gender-responsive pedagogy module of the EQUIP-T teacher in-service training being rolled out between midline and endline, in 2017. Furthermore, in only 2% of lessons observed at endline teachers used examples that challenged gender stereotyping, such as *father cleans house or boys wash plates*.⁷⁸

Table 7: Gender balance in teachers' interactions with pupils in the classroom (trends in programme areas)

	Baseline		Midline		Endline		Difference	
	Estimate	N	Estimate	N	Estimate	N	BL-EL	ML-EL
Teachers' interactions with pupils in the classroom is (% lessons observed)								
Gender-balanced	53.7	190	65.02	219	68.48	192	14.78**	3.46
More with boys	30.8	190	23.16	219	18.64	192	-12.16**	-4.52
More with girls	15.51	190	11.82	219	12.88	192	-2.63	1.06

Sources: Evaluation baseline, midline, and endline surveys (lesson observation).
Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (2) This is for all Standard 2 maths and Kiswahili observed lessons.

The change in the spatial inclusion of teachers' interactions with pupils in the classroom has been mixed. On one hand, in 76% of lessons observed, teachers were engaging with at least one pupil from all six areas of the classroom (Table 8) and this has increased significantly since baseline (by 17 percentage points), although there has been no significant change since midline. On the other hand, the proportion of teacher interactions with pupils in the back two areas of the classroom has significantly decreased since both baseline and midline. There is considerable scope for improvement as pupils in the front of the classroom still receive the most attention from teachers, while pupils in the back receive the least attention.

Table 8: Spatial inclusiveness of teachers' interactions with pupils in the classroom (trends in programme areas)

	Baseline		Midline		Endline		Difference	
	Estimate	N	Estimate	N	Estimate	N	BL-EL	ML-EL
Teacher interacts with at least one pupil from all six areas of the classroom (% lessons observed)	58.47	190	79.33	225	75.59	192	17.12***	-3.73

Distribution of teachers' interactions with pupils by classroom area (mean % teacher interactions)

⁷⁷ Refer to Volume II, Annex F, for a description of how this indicator is constructed.

⁷⁸ This indicator is part of EQUIP-T's logframe indicator for the use of gender-responsive pedagogy.

Front two areas	41.61	190	37.99	219	41.39	192	-0.22	3.39**
Middle two areas	30.53	190	34.42	219	34.39	192	3.86**	-0.02
Back two areas	27.86	190	27.59	219	24.22	192	-3.64**	-3.37***

Sources: Evaluation baseline, midline, and endline surveys (lesson observation).

Notes: (1) Asterisks indicate statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. (2) This is for all Standard 2 maths and Kiswahili observed lessons. (3) Note that the distribution of teachers' interactions is reported as the average proportion of teacher interactions across all lessons.

4.5.2 Have teaching and assessment practices improved? (EQUIP-T intermediate outcome)

Use of positive teaching practices in the classroom

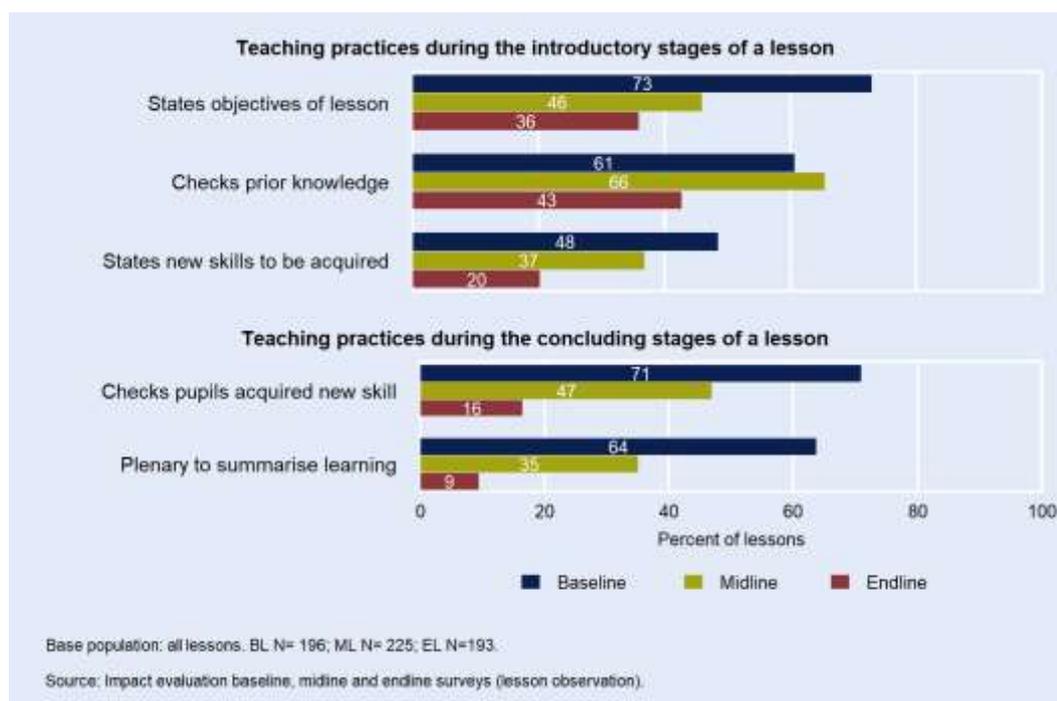
The evaluation measures 14 teaching practices that are considered to characterise positive teaching and classroom practices (Siraj *et al.*, 2014; Westbrook, 2013) at different stages in the lesson—the start, the middle, and the end.⁷⁹ These are not intended to be an exhaustive list of practices that can support pupil learning effectively, but rather give an indication of some of the core practices happening in classrooms in EQUIP-T schools over time.⁸⁰

It is important to note that the introduction of the 3Rs curriculum in 2015 reduces the comparability of the following findings with the baseline, whether between baseline and endline, or baseline and midline. This is because under the new curriculum, the lessons are organised differently, whereby two or three of the 3Rs subjects (reading, writing, and arithmetic) are taught consecutively, with no break. This leads to lessons flowing into each other, with less defined introductions and conclusions. The comparability of the results between midline and endline, however, has not been compromised as lessons were organised in the same way in these two survey rounds.

⁷⁹ The teaching practices descriptors are provided in Annex E and the definitions of the teaching practices indicators are provided in Volume II, Annex F.

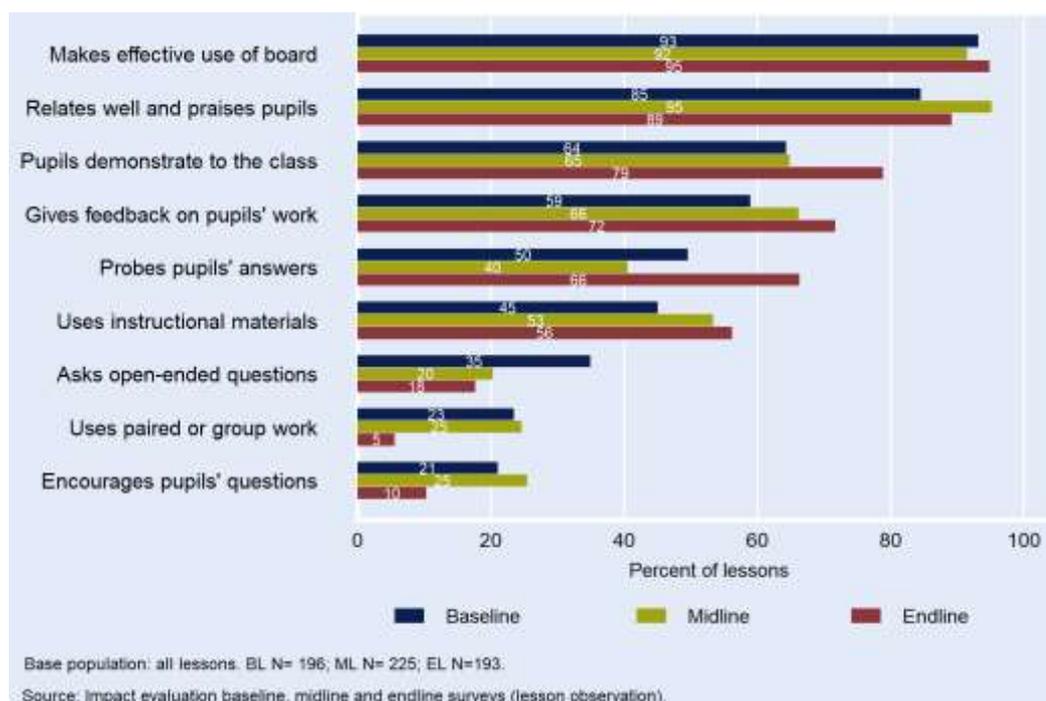
⁸⁰ EQUIP-T measures some similar indicators (as well as others) in its annual monitoring survey's lesson observation tool. This is used to report on its early primary teacher performance in pedagogy logframe indicator.

Figure 11: Observed teaching practices at the start and end of a lesson (trends in programme areas)



The use of positive teaching practices in the introductory lesson stages has significantly and substantially worsened since both baseline and midline (Figure 11). Teachers stated the objectives of the lesson and introduced the topic in a clear way in only 36% of the observed lessons, a drop of 37 percentage points since baseline, and 10 percentage points since midline. In only 20% of lessons did teachers state what new skills or knowledge pupils would acquire by the end of the lesson, down from 48% at baseline and 37% at midline. Teachers checked for prior knowledge of the lesson topic among the pupils in 43% of lessons, which has similarly declined considerably since baseline (61%) and midline (66%).

The significant decline in the use of positive teaching practices in the concluding stages of the lesson since baseline and midline is particularly striking. While a majority of teachers at baseline were demonstrating positive practices at the end of the lesson, a minority are doing so at endline. Teachers checked whether pupils have acquired the new intended skills or knowledge set out in the introduction in only 16% of lessons, a drop of 77% relative to baseline and 65% relative to midline. Similarly, the share of lessons in which teachers draw the whole class together at the end of the lesson to summarise what has been covered and to consolidate and extend learning has significantly declined, from 64% at baseline to 35% at midline, and finally to 9% at endline.

Figure 12: Observed teaching practices during the middle stages of a lesson (trends in programme areas)

Changes in the use of positive teaching practices during the middle stages of a lesson have been mixed (Figure 12).⁸¹ The biggest improvement has been in teachers probing or commenting on pupils' answers (66% of lessons at endline, up 17 percentage points from baseline and 26 percentage points from midline). The share of lessons where teachers ask pupils to demonstrate in front of the class has also significantly increased, from about 64% at baseline and midline to 79% at endline. Compared to baseline, more teachers at endline provide written or verbal feedback to pupils on their individual work (a significant increase from 59% to 72%). Use of different instructional materials, such as maps, posters, tables, charts, and real-life items, but excluding the blackboard, is also more common at endline (56%) than it was at baseline (45%). This is a positive finding given that EQUIP-T supported teachers during the in-service teaching training to use the available instructional materials and to develop home-made teaching aids for use in the classroom. Teachers make effective use of the chalk/blackboard in the overwhelming majority of lessons (95%), similarly high at baseline and midline.

On the other hand, a mere 5% of lessons at endline involve pupils carrying out activities in pairs or in groups, a significant drop of more than 75% relative to baseline and midline. The share of lessons in which teachers ask open-ended questions has also dropped by 17 percentage points since baseline, to reach only 18% at endline. Another deterioration has been in encouraging pupils to ask questions, which has declined significantly from baseline (21%) and midline (25%), to only 10% of lessons at endline. While teachers have a good rapport with pupils and are using praise in the majority of lessons at endline (89%), this has declined slightly since midline by 6 percentage points, but with no significant change since baseline. Additionally, the proportion of Kiswahili lessons in which teachers listen to most pupils individually reading a list of sounds, words or a paragraph during the lesson has declined significantly, from 51% at midline to 20% at endline.⁸²

⁸¹ The endline lesson observation instrument observed 11 teaching practices during the middle stages of the lesson, two of which are discussed elsewhere in this report. These are switching between Kiswahili and a vernacular language and using examples that challenge gender stereotyping (discussed in Section 4.5.1).

⁸² This practice was not measured at baseline.

In summary, a minority of teachers demonstrate a range of positive teaching practices in the classroom and this has significantly worsened over time. At baseline, teachers were demonstrating at least seven positive teaching practices, out of the measured 14 practices,⁸³ in the majority of lessons (68%) but that significantly declined to 56% of lessons at midline and has decline further to 40% at endline.⁸⁴ This suggests that there is considerable scope for the further strengthening of general pedagogy.

The decline in the use of many of the positive teaching practices measured in this evaluation relative to both baseline and midline is likely to be the result of a number of factors. The first is the considerable increase in class sizes in the programme schools since baseline and midline. In observed lessons with more than 80 pupils, teachers were less likely to use paired or group work, use instructional aids, and to use at least seven positive teaching practices. That said, there were several teaching practices that improved over time despite this massive increase in class sizes, including probing or commenting on pupils' answers, asking pupils to demonstrate in front of the class, and providing written or verbal feedback to pupils on their individual work. These are very positive findings given that all of these practices can be considerably constrained in large classes. Another factor could be the introduction of the new 3Rs curriculum, which led to a reorganisation of lessons between baseline and endline, with lessons now taught consecutively and flowing into each other. It is also possible that the EQUIP-T training led to the improvement of other teaching practices not captured in this evaluation; however, the decline in many of the observed practices is a concern.

A written lesson plan was available and was seen by an enumerator in less than half of the lessons observed (47%) and this has significantly worsened since baseline (66%) and midline (63%). A possible reason for this deterioration might be the increase in workload. Teachers at endline are preparing and marking more pupil assessments (discussed in the next sub-section) and for much larger class sizes, and they are also participating in more meetings due to the introduction of the COL, and as a result this may lead to less time for lesson planning. Regardless of the reason, this is a concern, given that lesson plans are an integral part of the teaching process.⁸⁵

Use of continuous pupil assessment

Use of continuous pupil assessment is a central aspect of the teaching and learning process. The 3Rs curriculum emphasises the use of assessments by teachers, including formal and informal continuous assessments based on pupils' daily work, as well as the provision of feedback to pupils and their parents on the pupils' academic performance (Ministry of Education and Vocational Training (MOEVT), 2016).

There has been some improvement in the use of regular pupil assessments by teachers.

While almost all Standards 1 to 3 teachers at endline (96%) report that they had assessed pupils' academic progress in the past five days, only 64% are able to provide a marked example of any assessment, and these shares have not significantly changed since baseline or midline. The most common types of assessment used by teachers at endline are class exercises (59%), followed by written class tests (26%) – the latter have dropped by 13 percentage points since baseline. On the

⁸³ 16 different teaching practices were recorded during the lesson observations, but the range of positive practices is based on a total of 14 practices, which exclude language switching and the use of examples to challenge gender stereotyping.

⁸⁴ This is a subjective benchmark intended to summarise trends in overall practices. It assumes that as the practices are fairly general (for example, encouraging pupils to ask questions), it is a reasonable benchmark to expect teachers to demonstrate half of these during a lesson.

⁸⁵ The evaluation does not measure whether teachers planned for a lesson without recording it on a lesson plan.

other hand, significantly more teachers at endline (14%) are able to show that they had assigned and marked homework in the preceding five days, relative to both baseline (5%) and midline (6%). However, that proportion is still at a very low level.⁸⁶ Furthermore, there has been an increase in the number of assessments used, as significantly more teachers at endline (36%) are able to provide a marked example of two or more types of assessments conducted in the past five days, a gain of 9 percentage points since baseline and 15 percentage points since midline.

There has also been an improvement in the share of teachers who are providing feedback to pupils and their parents on the pupils' academic progress. Among all Standards 1 to 3 teachers at endline, 83% state that they report individually on their pupils' academic progress to the pupils' parents or guardians. On average, teachers at endline report providing feedback to parents 2.5 times in the previous school year, a significant improvement since baseline. On the other hand, interviewed parents report that, on average, they received written information from the school or met with a teacher to discuss their children's academic progress 0.9 times in the last school year. Furthermore, the share of lessons where teachers provide feedback to the pupils directly on their individual work has increased significantly since baseline, from 59% to 72%.

The improvements in the use of continuous pupil assessment and provision of feedback to pupils and parents are positive findings, especially in light of the substantial increases in class sizes.

4.6 Findings: provision, availability, and use of EQUIP-T TLMs

Provision of TLMs is a key input of component 1A, which EQUIP-T expects will lead to increased availability of TLMs in classrooms and ultimately to their increased usage by teachers (Figure 7).

4.6.1 Have schools received EQUIP-T TLMs? (EQUIP-T input)

EQUIP-T TLMs have generally reached schools, but not all schools have received all of the materials. While 97% of schools report receiving 'big books' since baseline, and 90% report receiving literacy toolkits and 88% report receiving supplementary readers, only 71% of schools report receiving the numeracy toolkits and only 36% of schools say they have received the teacher 'read-aloud' books.

4.6.2 Has the availability of TLMs in classrooms increased? (EQUIP-T output)

The availability of EQUIP-T provided TLMs in the classrooms is low. Despite the majority of schools reporting receiving the supplementary readers, these readers were available in the classrooms in only 14% of observed Standard 2 Kiswahili lessons (unchanged since midline).

The availability of other TLMs in the classrooms, which are not directly provided by EQUIP-T, is generally higher. Instructional materials were displayed on the walls in 94% of lessons observed, up 49 percentage points from baseline and 8 percentage points from midline. This is a positive finding given that EQUIP-T supported teachers during the in-service teaching training to develop home-made teaching aids for use in the classroom. The majority of pupils (92%) in an observed lesson had a pencil, while 94% had a maths exercise book, and 79% had a Kiswahili exercise book.

⁸⁶ This is in line with 19% of Standard 3 pupils reporting that they had been given homework in the past five days.

In the maths lessons observed, the pupil–maths textbook ratio was 2.5:1, while in the Kiswahili lessons observed, the ratio was 3.2:1.⁸⁷ While these ratios are high, they have decreased significantly since midline and are in fact lower than the Government ratio target of 4:1 set for the national EPforR programme (MOEST and PO-RALG, 2018).

4.6.3 Has the use of TLMs in classrooms increased? (EQUIP-T intermediate outcome)

The use of EQUIP-T-provided TLMs in the classrooms is limited. Despite the majority of schools receiving ‘big books’ or ‘read-aloud’ books, these were used in only 4% of Kiswahili lessons observed, with a similarly low level at midline. In 40% of maths lessons observed, no maths learning materials were used by any of the pupils. In the very few Kiswahili lessons observed where the supplementary readers were available in the classroom, none of the pupils read them during the lesson.

Use of other TLMs is also low. Textbooks were used by pupils in only 28% of the lessons observed, up 15 percentage points from midline but down 10 percentage points since baseline. While the share of observed lessons in which teachers use any instructional materials has increased since baseline (as shown in Section 4.5.2), most likely owing to the presence of more teaching aids in the classrooms compared to baseline, and to EQUIP-T’s training on the use of TLMs, there remained a sizeable minority of lessons (44%) in which no materials were used at all.

The limited use of EQUIP-T materials in classrooms was also a strong finding from the midline evaluation. That this situation has changed so little by endline seriously compromises the intended effect of the EQUIP-T materials on improving pupil learning. It is well known that the availability and use of instructional material and textbooks has been strongly associated with improving education quality in many contexts (UN Educational, Scientific and Cultural Organization (UNESCO), 2015).

4.7 Findings: instructional time and teacher attendance

The amount of instructional time that pupils receive directly influences the amount of learning that can take place, and the loss and inefficient use of instructional time has been shown to be linked with poor education quality (Moore, DeStefano, and Adelman, 2012; Robinson, Lloyd, and Rowe, 2008; UNESCO, 2015). Having appropriate instructional hours for early grade standards is one of the intermediate outcomes in the theory of change of component 1A (Figure 7).

Teacher absenteeism takes a toll on pupils’ learning as it causes substantial losses in instructional time. EQUIP-T seeks to implicitly improve teachers’ motivation, in part to reduce their absenteeism and thereby increase their time on task (Cambridge Education, 2014; EQUIP-T MA, 2015).

4.7.1 Teacher absence from school and classrooms, and punctuality⁸⁸

Teachers’ absence from school has not changed significantly since baseline or midline. Of all teachers at endline, 13% were absent from school on the day of the survey, and this is largely unchanged since baseline and midline (Table 9). The impact estimation finds some evidence that

⁸⁷ This is supported by findings from the teacher interviews, where 47% to 49% of Standards 1 to 3 teachers reported having limited to no access to maths and Kiswahili textbooks, respectively, for the majority of their Standard 2 pupils.

⁸⁸ These indicators are measured using the headcount instrument. Refer to Volume II Annex F for details on the measurement.

EQUIP-T has reduced teachers' school absenteeism between baseline and endline, but the results are only weakly significant (see Box 9).

Table 9: Teacher absenteeism and punctuality on the day of the survey (trends in programme areas)

	Baseline		Midline		Endline		Difference	
	Estimate	N	Estimate	N	Estimate	N	BL-EL	ML-EL
On the day of the survey, of all teachers in the roster:								
Absent from school (%)	12.05	1,005	13.53	1,074	13.48	1,014	1.43	-0.05
Of all teachers present on the day of the survey:								
Arrived late (%)	63.23	873	55.42	923	40.2	862	23.03***	-15.22***
Of all teachers present on the day of the survey and timetabled to teach:								
Absent from classroom (%)	66.79	708	61.06	675	60.7	396	-6.09	-0.36
Of all Standards 1 and 2 teachers present on the day of the survey and timetabled to teach:								
Absent from classroom (%)	57.64	144	36.26	150	34.53	85	23.11***	-1.74
Sources: Evaluation baseline, midline, and endline surveys (teacher head count).								
Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (2) Mean over all teachers.								

There has been a significant improvement, however, in teachers' punctuality. While the share of teachers arriving late to school is still quite high (40%), it has significantly decreased, by 23 percentage points, since baseline, and by 15 percentage points since midline.

The trend in teachers' absenteeism from the classroom differs for early grade and upper grade teachers. The classroom absenteeism rate for all teachers at endline is 61% – that is, 61% of teachers are not present in the classrooms despite being present at school and timetabled to teach.⁸⁹ This is an extremely high share and remains largely unchanged since baseline and midline. The impact estimation finds some evidence that EQUIP-T has reduced teachers' classroom absenteeism between baseline and endline, but the results are only weakly significant (see Box 9). On the other hand, the classroom absenteeism rate for Standards 1 and 2 teachers is much lower, at 35%, and this has dropped significantly, by 40%, since baseline. There has been no significant change since midline, however.⁹⁰ The qualitative research at midline found that EQUIP-T was perceived to have had a positive effect on early grade teachers' motivation, which could potentially have led to lower classroom absenteeism for Standards 1 and 2 teachers (OPM, 2017a).

Despite the lower classroom absenteeism rate for Standards 1 and 2 teachers, and its improvement since baseline, a rate of 35% is still quite high, and leads to large losses in instructional time (see Section 4.7.2), severely reduces the time during which teachers use and further develop teaching practices learnt during their in-service training, and consequently forms an important barrier to improving pupils' learning. On top of that, when teachers do attend their lessons, they might be away for parts of the lesson. At endline, teachers left the classroom in 24%

⁸⁹ About 8% of teachers at endline were absent from classrooms on the day of the survey at the time of the headcount observation because they were engaging with the survey team. Adjusting for this, classroom absenteeism for all teachers would be 59%. The share of absent teachers at baseline who were engaged with the survey team was not measured, and therefore it is not possible to compute a classroom absenteeism indicator adjusting for that, but it is expected to be similar to endline – that is, the change would be downward but minimal.

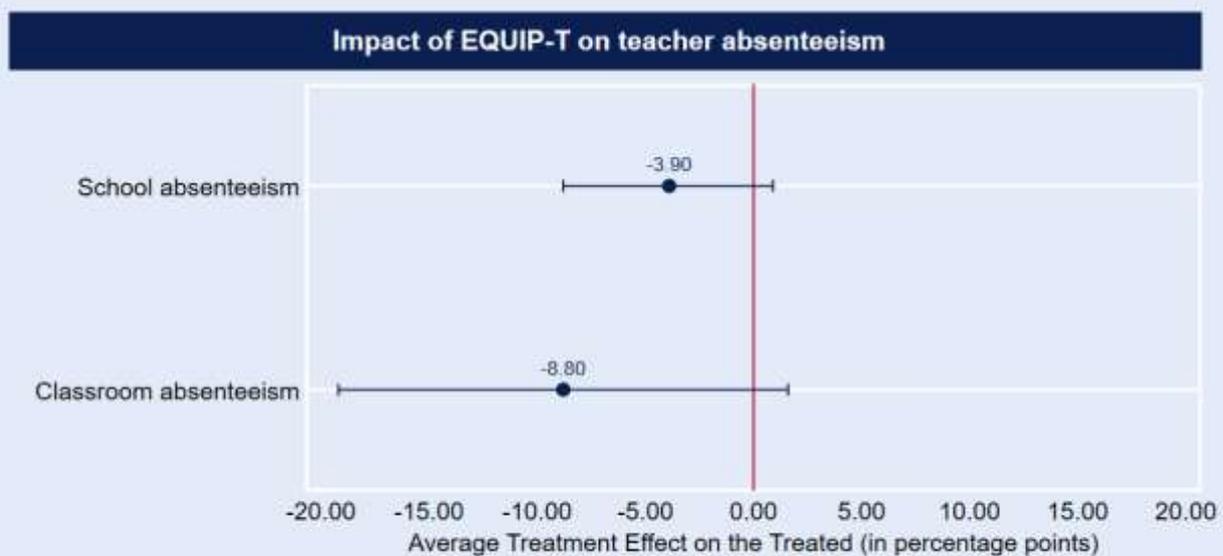
⁹⁰ The impact estimation does not measure EQUIP-T's impact on absenteeism indicators for Standards 1 and 2 teachers because of sample size limitations.

of lessons observed and were away for 3.5 minutes on average.⁹¹ This further compounds the classroom absenteeism rate and leads to further reduction in instructional time.

Large workload is the most common reason reported by teachers for absence from the classroom (31% of teachers). This is followed by meeting with other teachers (28%); illness (27%); and meeting with the head teacher (14%). At endline, teachers have much larger class sizes, which increases the volume of marking, and they are also expected to attend more meetings as a result of the COL approach. This means that EQUIP-T may unintentionally be contributing to factors which might in turn contribute to classroom absenteeism. Head teachers views on reasons for teachers' classroom absence are in Volume II (Chapter 5).

Box 9: EQUIP-T's impact on teachers' absence from school and classrooms

The figure below shows the ATT detected on the proportion of teachers absent from school and skipping class (if timetabled to teach) on the day of the survey (in percentage points). It compares changes in EQUIP-T schools to changes in control schools between baseline and endline.



Source: IE baseline and endline surveys.

Note: Point estimates correspond to PSM with Difference-in-Difference estimates over baseline and endline survey data, 95% confidence intervals plotted.

There is some evidence that EQUIP-T has reduced teachers' school absenteeism, but the results are only weakly significant. The teachers in programme schools are found to be around four percentage points less likely to be absent from school because of EQUIP-T. Given the descriptive results presented in Table 9, it can be inferred that in the absence of the EQUIP-T intervention (the counterfactual situation) the proportion of teachers absent from school would have been around 17%, rather than the measured 13.5%. The result is weakly significant in statistical terms (around 10% significance), and the confidence intervals of the impact estimate slightly overlap with zero, as illustrated in the figure above. The sample of teachers used for the main estimation of impact is sizeable. However, the sample is considerably reduced in the robustness check model, in which the positive result on school absenteeism is strengthened both in terms of magnitude and statistical significance. Overall, the evidence of programme impact cannot be considered strong, but the fact that robustness checks consistently support this finding makes it stronger. This indication of impact is also more robust than the midline result, which showed no consistent evidence of impact across different estimation models.

There is some evidence that EQUIP-T has reduced teachers' classroom absenteeism, but the results are only weakly significant. The impact estimates shown in the figure above indicate that

⁹¹ The reasons for leaving the classroom were not recorded.

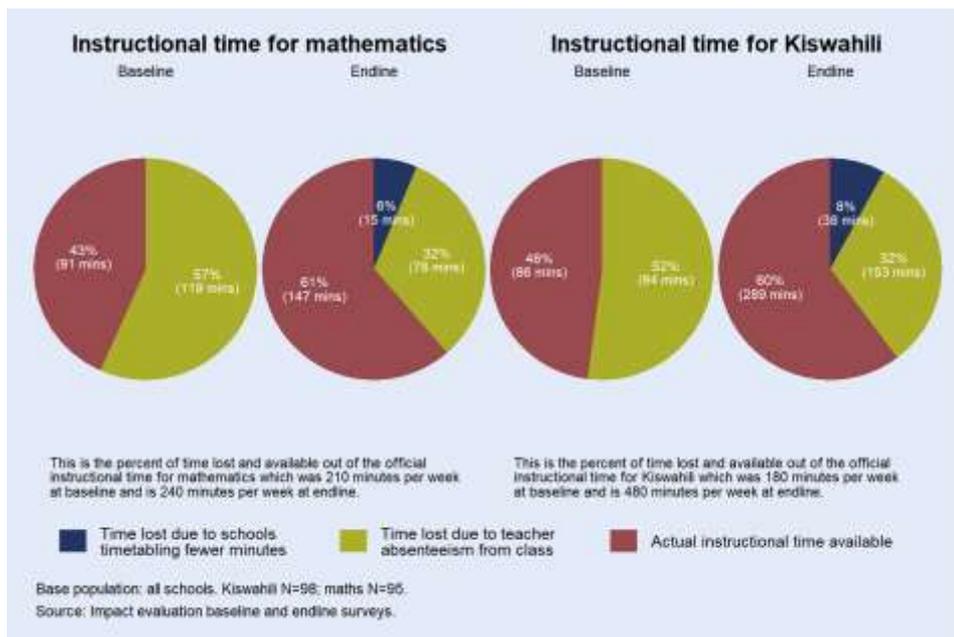
teachers in programme schools are around nine percentage points less likely to be absent from the classroom because of EQUIP-T. This implies that in the absence of EQUIP-T (the counterfactual situation) the percentage of teachers skipping classes would have been close to 70%, rather than the 60% reported in Table 9. However, the significance of this result is low in statistical terms (around 10% significance). The sample of teachers used for this estimation is small, mainly due to a low proportion of sampled teachers being timetabled to teach during the period observed on the day of the endline survey. This has a particularly detrimental effect on the robustness check strategy, which does not perform well and cannot confirm the results of the main estimation strategy. For these reasons, the robustness of the evidence of programme impact on this indicator cannot be considered strong. The endline evidence on classroom absenteeism is also weaker than the midline finding, which clearly pointed towards a positive and highly significant programme impact on this measure of absenteeism. However, it is worth bearing in mind that the sample size on which the endline impact estimation was performed was reduced compared to midline, due to the lower proportion of teachers timetabled to teach at endline. This may help explain why the evidence of impact detected at endline is not as strong as at midline.

4.7.2 Has instructional time increased? (EQUIP-T intermediate outcome)

The new 3Rs curriculum increased the official instructional time for Kiswahili for Standards 1 and 2 from 180 minutes to 480 minutes per week, and for maths from 210 to 240 minutes per week.⁹²

As at baseline and midline, there remains a substantial loss in actual instructional time, compared to the time intended. The actual amount of learning time that early grade pupils are getting is much lower than the official intended instructional time (Figure 13). One of the reasons for this is that head teachers are scheduling fewer instructional hours than the official guidelines—a practice that was not evident at baseline. Only 48% of schools at endline schedule the official required 480 minutes per week for Kiswahili and 63% schedule the official 240 minutes per week for maths. This may in part be explained by the double-shifting of classes which tends to squeeze available instructional hours. Due to large increases in enrolment coupled with teacher and classroom shortages, schools are increasingly having two shifts of classes. At baseline, 48% of schools were double-shifting classes compared to 73% of schools at endline (see Section 5.6.1).

Figure 13: Actual instructional time for Kiswahili and mathematics for Standards 1 and 2 (trends in programme areas)



⁹² The official requirements for the old 2005 curriculum were provided by the Director of Primary Education, as per the Education Circular of 2001. The requirements for the new curriculum come from MOEVT (2016).

The main factor for the loss of instructional time, however, is the high share of teachers not attending their lessons. If an adjustment to the timetabled weekly minutes is made to account for the classroom absenteeism rate of Standards 1 and 2 teachers in the programme schools (35%), pupils are receiving, on average, only 289 minutes of Kiswahili and 147 minutes of maths lessons per week.⁹³ This is a loss of 40% of instructional time for Kiswahili and 39% for maths, which likely has adverse effects on pupil learning. On top of these very large losses, there is likely to be even more loss when teacher absence from school, teacher late arrival, pupil absenteeism,⁹⁴ and inefficient use of lesson time are taken into account (for example, teachers spending some of the lesson time marking pupils' work as opposed to teaching).

Actual instructional time has increased dramatically between baseline and endline.

Compared with baseline, pupils are receiving 56 more minutes (maths) and 203 more minutes (Kiswahili) of instructional times per week on average. This is largely due to the introduction of the 3Rs curriculum, but also partly because the percentage loss in instructional time fell. As can be seen from Figure 13, the percentage of loss in instructional time for maths has declined from 57% at baseline to 38% at endline, and similarly the loss in instructional time for Kiswahili has declined from 52% to 40%. This is mainly due to the significant decrease in the classroom absenteeism rate for Standards 1 and 2 teachers between baseline and endline. Actual instructional time remained the same between midline and endline, largely because classroom absenteeism rates did not change.

4.8 Summary of findings on teacher performance

This section summarises the evaluation evidence on EQUIP-T component 1A: teacher performance. Table 10 provides an overview of the extent to which EQUIP-T has achieved its intended inputs, outputs, and intermediate outcomes, and the factors that may have influenced their achievement.

Table 10: Summary of evidence on teacher performance

Evaluation questions and assumptions related to the theory of change	Result at EL	Significant change	
		BL to EL	ML to EL
Has teacher performance improved in treatment schools, with changes consistent with the theory of change?			
Inputs Has EQUIP-T provided teacher in-service training and TLMs?			
Teacher attended EQUIP-T in-service training in last two years (% Stds. 1 and 2 teachers)	98%	..	none
School held 10 or more school-based in-service training sessions last year (% schools)	52%	..	positive
Teacher completed all Kiswahili, numeracy, and gender-responsive pedagogy modules (% Stds. 1 and 2 teachers)	35%
School received numeracy teaching aid toolkit (% schools)	71%
Outputs Has teacher capacity and confidence increased, are TLMs available in classrooms, and are COL operating?			
Teacher confidence in teaching the new 3Rs curriculum is very high (% Stds. 1 and 2 teachers)	78%	..	none
At least one supplementary reading book available in classroom (% Kiswahili lessons)	14%	..	none
School participated in at least one ward cluster reflection meeting in 2016–17 (% schools)	77%
Teacher attended four or more SPMs in last 60 days (% Stds. 1 to 3 teachers)	55%

⁹³ Refer to Volume II, Annex F, for the definition of the actual instructional time indicators.

⁹⁴ 28% of Standards 1 to 3 pupils were absent from school on the day of the survey (Section 3.2.2 in Chapter 3).

TOC assumptions What factors may have influenced the delivery of outputs?			
Teacher attended all school-based in-service training sessions (% Stds. 1 and 2 teachers)	52%	..	negative
Teacher does not report difficulties with EQUIP-T in-service training (% Stds. 1 and 2 teachers)	23%	..	negative
Intermediate outcomes Has classroom teaching improved and are instructional hours appropriate?			
Teacher interactions with pupils in the classroom are gender-balanced (% lessons)	68%	positive	none
Teacher interacts with at least one pupil from all areas of the classroom (% lessons)	76%	positive	none
Teacher uses at least seven positive teaching practices in the classroom (% lessons)	40%	negative	negative
Teacher uses different instructional materials during the lesson (% lessons)	56%	positive	none
Teacher uses 'big books' or 'read-aloud' books during lessons (% Kiswahili lessons)	4%	..	none
Most pupils use maths learning materials during the lesson (% maths lessons)	39%	-	-
Teacher shows two or more types of pupil assessments conducted in the past five days (% Stds. 1 to 3 teachers)	36%	positive	positive
Instructional time timetabled for Stds. 1 and 2 maths meets official requirements (% schools)	63%	none	none
Instructional time timetabled for Stds. 1 and 2 Kiswahili meets official requirements (% schools)	48%	negative	none
TOC assumptions What factors may have influenced the delivery of intermediate outcomes?			
Average class size for primary standards is 60 pupils or less (% schools)	31%	negative	negative
Teacher still at the same school since baseline (% all baseline teachers)	50%
Teacher present at school on the day of the survey (% all teachers)	87%	none	none
Teacher present in classroom when in school and scheduled to teach (% all teachers)	39%	none	none
Teacher present in classroom when in school and scheduled to teach (% Stds. 1 and 2 teachers)	65%	positive	none
Teacher arrives to school on time (% all teachers)	60%	positive	positive
Teachers' job satisfaction is higher compared to two years ago (% Stds. 1 to 3 teachers)	36%	-	negative
Notes: (1) .. means not applicable either because it is an EQUIP-T input, or because of the nature of the indicator. (2) - means trend data are not available.			

While the residential component of the EQUIP-T in-service training has been delivered largely as planned, there are gaps in the implementation of the school-based component.

Nearly all Standard 1 and 2 teachers have received EQUIP-T training in the last two years; however, there are still some teachers who have not attended any residential training and others that have not attended any school-based training, despite teachers being expected to attend at least some residential training and all of the school-based sessions. The school-based training, which is at the core of the in-service training model, is not being held as regularly as intended in the majority of schools, despite its improvements over time. The majority of school-based sessions cover Kiswahili literacy and numeracy modules, and very few cover particular curriculum competencies, which is at odds with the in-service training model's evolution into a COL approach. Early grade teachers' regular attendance at the school-based sessions has also been worsening over time.

The result is that a very small share of teachers have completed all the in-service training modules on early grade Kiswahili literacy, early grade numeracy, and gender-responsive pedagogy. These training shortfalls will weaken any effect of EQUIP-T on teacher performance expected to be achieved through the training. The main challenges reported with the school-based in-service training are no allowance, participants not being motivated, and sessions taking place at inconvenient times. On the other hand, there have been large spill-overs, with many teachers of Standards 4 to 7 also receiving the in-service training both away from school and in school.

Standards 1 and 2 teachers' confidence in teaching the new 3Rs curriculum remains high, and COL structures are operating – but not as frequently as intended. While the majority of

schools are participating in the ward cluster reflection meetings, a sizeable minority are not doing so. Among schools that are attending these meetings, attendance is much lower than the expected quarterly frequency. SPMMs are taking place in most schools but not always on a weekly basis and for their intended purpose.

Teachers' performance in the classroom has improved in some respects but worsened in others. Teachers' interactions with pupils in the classroom have become significantly more gender-balanced and to some degree more inclusive of pupils seated in different parts of the classroom. However, interaction with pupils seated at the back of the classroom has declined. Only a minority of teachers in the observed lessons used a range of positive teaching practices and this share has declined over time. Lesson planning has also fallen over time. On the other hand, there has been positive change in some teaching practices, an increase in the number of methods used by teachers to assess pupils' academic progress, and teachers have been providing more feedback to pupils and their parents.

While the majority of schools report receiving TLMs from EQUIP-T, the availability and use of these materials in the classroom is very limited.

The difference between intended and actual instructional time for Standards 1 and 2 has declined considerably since baseline but still remains very high. This is mostly due to the high rate of teacher classroom absenteeism. There has been a significant and large decrease of this rate for Standards 1 and 2 teachers, and there is some evidence that EQUIP-T had a positive impact on the classroom absenteeism rate for all teachers between baseline and endline.⁹⁵ Nonetheless, classroom absenteeism remains very prevalent and this reduces the intended effect of EQUIP-T in-service training on teacher effectiveness.

The steep growth in class sizes since baseline and midline, and the high level of teacher and INCO turnover, have made achieving the anticipated changes in teaching practices and instructional time much more challenging. Teachers' workloads have increased, and although they report high levels of job satisfaction, satisfaction compared to two years ago has deteriorated. Applying some of the teaching methods learnt during training becomes far more difficult when classrooms are seriously overcrowded. Nonetheless, these conditions were already unfavourable at midline, and were reported in the midline evaluation as posing risks to programme impact and sustainability.

⁹⁵ The impact detected was weakly significant.

5 SLM

The activities under EQUIP-T component 2 aim to strengthen SLM to promote whole-school development (EQUIP-T MA, 2017). This chapter begins by providing an overview of the results chain for SLM from the programme theory of change,⁹⁶ programme implementation since baseline, and changes expected by the programme as a result of its activities. This is followed by the quantitative findings on SLM, which are structured according to the results chain. Finally, the chapter summarises key findings, and maps out the extent to which different inputs have been successful in achieving their intended outputs and intermediate outcomes, as well as factors that may have affected their ability to achieve them.

The chapter also covers some outputs and assumptions concerning WEOs from EQUIP-T component 3B, district planning and management.

Box 10 describes some of the key SLM concepts for EQUIP-T.

Box 10: Description of key EQUIP-T SLM concepts

PTP: Class-based group of parents and teachers. The focus is on classrooms, with parents actively supporting classroom activities and helping to address specific problems, such as pupil absence (see Chapter 6 on EQUIP-T component 4A for more information).

School development plan (SDP): A plan that sets out the school's mission and vision statements and context (academic, infrastructure, and financial), that contains an action plan with measures to improve school performance based on internal assessment, taking into account school needs, priorities, and resources.

SIS: A management tool used for education planning, monitoring, and decision making, based on data collected at school level. Under EQUIP-T, schools are meant to use the SIS to improve their performance by increasing the use of data for decision making and oversight.⁹⁷

5.1 SLM: theory of change, implementation, and expectations of change

5.1.1 Theory of change

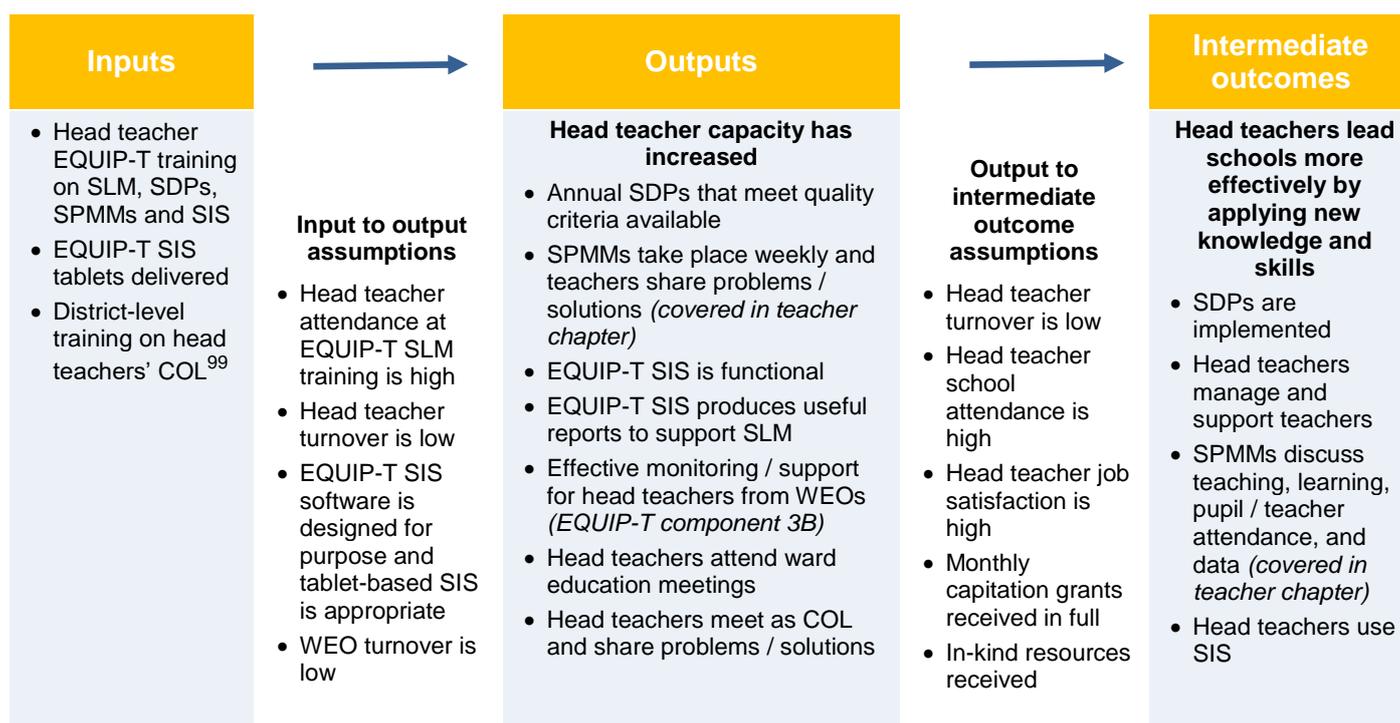
Improved SLM supported by more effective monitoring and support from WEOs is a key part of the model for whole-school improvement embedded in the programme theory of change (OPM, 2018). Figure 14 shows the inputs, outputs, and intermediate outcomes and assumptions in the main results chain for EQUIP-T component 2, SLM, which are covered by the quantitative research.⁹⁸

⁹⁶ The SLM results chain presented here shows aspects of the theory of change addressed by the quantitative research.

⁹⁷ SIS data are also used at central, regional, and district levels.

⁹⁸ WEO training, grants, and transport are not examined in this report.

Figure 14: Main results chain for EQUIP-T component 2, SLM



5.1.2 Implementation since baseline

The main activities between baseline and midline were training for head teachers, assistant head teachers, and WEOs on EQUIP-T SLM modules 1 and 2. The first module covered education quality standards and introduced the National School Leadership Competency Framework, covering the responsibilities and roles of head teachers. The second module aimed to prepare head teachers and WEOs to lead EQUIP-T school-level initiatives, and topics covered included PTP grant management, SRP set-up, SIS introduction, and school development activities, including communication. Training on SLM module 3 on school development planning was also started (OPM, 2016a).

Since midline, three main activities were delivered: the training for head teachers and WEOs on school development planning was completed and there was also additional support to review draft SDPs; training took place for head teachers and WEOs on holding SPMMs; and the provision of tablets to all head teachers was accompanied by initial training on the SIS (OPM, 2018).

Annex B.3 provides an overview of SLM activities implemented since baseline to achieve the EQUIP-T aim of strengthened SLM.

5.1.3 Expected changes

By midline, the programme's expected changes at school level were head teachers being more aware of their responsibilities and roles; head teachers being more transparent, especially about school finances; head teachers using the school notice board to communicate (for results on this, see Section 6.6 in Chapter 6); head teachers keeping better school records and improved record

⁹⁹ This is training of trainers for district officials on the school leadership COL initiative for head teachers. The formal training of head teachers had not yet taken place by the time of the endline survey, but informal information sharing between district officials, WEOs and head teachers may have taken place.

management; head teachers having better relationships with teachers and WEOs; and schools running more extra-curricular activities (OPM, 2016a).

Between midline and endline the changes expected by EQUIP-T staff were: improved attitude of head teachers and willingness to reform; head teachers with strong leadership skills; SPMMs taking place weekly, with minutes available; SDPs available, containing activities to promote inclusion, and being implemented; greater financial transparency, including noticeboards with financial information displayed; and proper fund management. Under EQUIP-T component 3B, WEOs were expected to be more aware of their professional responsibilities and to conduct more school visits (OPM, 2018).

The EQUIP-T logframe also includes selected indicators to capture expected changes under component 2. These cover: head teacher performance management;¹⁰⁰ SDP implementation; upload of pupil data to the SIS; and whether wards have active SLM COL for head teachers. For more details on logframe indicators see Annex B.6.

5.2 Findings: head teacher background and turnover

In each of the findings sections, quantitative evidence from the evaluation's baseline, midline, and endline surveys is presented to assess if changes in SLM have occurred as expected over time. Selected key assumptions in the SLM results chain are also explored, starting with head teacher characteristics and turnover.

5.2.1 Head teacher background characteristics and turnover

At endline, the proportion of female head teachers remains small, at 16%, and head teachers are on average significantly younger than at baseline, by three and a half years, with an average age of 40 years. Head teachers have similar academic qualifications at endline to those they had at baseline. The vast majority (80%) have completed Form 4, 14% have a Form 6 qualification, and the remainder have either a Certificate or a Diploma / Advanced Diploma in Education.

However, for professional qualifications there have been significant changes over time. At endline, 5% of head teachers have a Bachelor's of Education degree or higher, whereas none did at baseline; 45% have a Diploma or Advanced Diploma in Education, compared to only 9% at baseline; and 50% have a Certificate in Education, compared to 90% at baseline. This change is largely driven by the new national policy implemented in 2017 that raised the minimum qualification requirement for head teachers to a diploma, while previously head teachers were appointed based on their efficiency and work experience, and a certificate was sufficient.

Box 11: Level of head teacher turnover¹ (EQUIP-T input to output and output to intermediate outcome assumption)

Only 26% of head teachers who were at the school at baseline are still at that same school by the endline four years later.

Head teacher turnover in the programme schools has been extremely high: 46% between baseline and midline, and 51% between midline and endline. There are two main reasons for the high turnover between

¹⁰⁰ The 10 head teacher performance management indicators are: school plan completed, with gender-responsive strategies; school plan implementation started and progress tracked; weekly management meetings with teachers; in-service training sessions at least once a month (minimum six sessions per term); school cash book in place and updated each quarter; school committees elected and meeting at least quarterly; school club in place and active; PTP in place and active; SIS data updated (there are both male and female data); and follow-up on teacher attendance from SIS data.

midline and endline: head teachers leaving the school (74%) or being demoted in the same school to assistant head teachers, academic masters, or teachers (26%), under the professionalisation drive for head teachers that started after midline. Among the head teachers who have left, by far the most common reason was being transferred (69%), while the remaining group of head teachers who have left retired (19%), died (6%), went away for studies (4%), or were seconded (1%).

At endline, among the head teachers who have been head teachers at their current school for less than two years, the majority (91%) came from another school in the same district. Because EQUIP-T covers all schools within the programme districts, this will to some degree alleviate the effect of head teacher turnover on SLM training if head teachers new to the school were also head teachers at their previous school and received SLM training, as they will continue to work in programme schools.

For the 8% of these head teachers who were teachers at the same school before being promoted, EQUIP-T's impact on SLM will be undermined as they will not have received the SLM training provided prior to the midline, and may only have received some of the SLM training provided since midline, depending on when exactly they were promoted to head teachers.

The professionalisation of head teachers is a one-off event that will not affect head teacher turnover going forward. However, the high level of head teacher transfers, the main contributor to the high turnover, is a persistent feature of the education system, and will continue to affect interventions aimed at head teachers.

Source: Evaluation baseline, midline and endline surveys.

Note: (1) Head teacher turnover is defined as the rate in which head teachers leave their current posts that are defined by their role (head teacher) and place of work. See Volume II Chapter 5 Section 5.4 for more details.

The average head teacher has worked as a head teacher for six years and has been head teacher at their current school for only three years. This is largely the result of the high head teacher turnover in the programme schools covered by the evaluation (see Box 11). More details on head teacher turnover can be found in Volume II Chapter 5 Section 5.4.

5.3 Findings: EQUIP-T training for head teachers

5.3.1 Has EQUIP-T provided SLM in-service training for head teachers? (EQUIP-T input)

The provision of SLM in-service training for head teachers has increased significantly since baseline. At baseline only 11% of head teachers had attended some type of SLM training during the last two years; by midline this had risen to 68%, and at endline it is 76%. The main SLM training provider at endline by far is EQUIP-T (76%), followed by BRN-Ed (3%), and STEP (2%).¹⁰¹

The EQUIP-T SLM training has not reached all head teachers, which will likely undermine EQUIP-T's impact on SLM. At endline 80% of current head teachers have **ever** attended EQUIP-T SLM training.

Between baseline and midline, 67% of head teachers attended EQUIP-T SLM training, meaning they should have received training on SLM modules 1,2 and 3 (see Section 5.1.2 and Annex B.5).¹⁰² There is a mitigating factor in that assistant head teachers were also meant to be trained. If they were, they could share the knowledge and skills acquired during the training with head teachers, although the effect of the SLM training in this case would likely be diluted, compared to if head teachers themselves had attended.

Between midline and endline, 76% of head teachers attended SLM EQUIP-T training, meaning they should have received training on school performance management, review of SDPs, and SIS

¹⁰¹ This is for all head teachers and among these 24% did not attend any SLM in-service training at endline.

¹⁰² Assuming they attended all training events.

management (phase 1).¹⁰² Among the 24% of head teachers who did not receive SLM training, the reason for all but one was that they had joined the school since the end of 2017, after the training had been conducted. Since EQUIP-T covers all schools within the programme districts, and the majority of head teachers who were new to their school both at midline and endline came from another school in the same district, they should have received EQUIP-T training if they were also head teachers at their previous school (see Box 11). At endline 30% of the head teachers who were new to the school worked as head teachers at their previous school and should have received EQUIP-T SLM training but did not, whereas 67% of the new head teachers worked as teachers at their previous school and 2% had another job in education which means they would not have received the SLM training. Examining EQUIP-T training coverage taking head teachers' turnover status into account yields further insight.

Table 11: EQUIP-T SLM in-service training received, by head teachers' turnover status (trends in programme areas)

Indicator	Midline		Endline		Difference
	Estimate	N	Estimate	N	ML-EL
Head teachers received EQUIP-T SLM training last two years (%)					
At school at endline only			54.93	51	
Same school at midline and endline	65.19	44	98	46	32.82***
Same school at baseline, midline and endline	70.32	26	96.14	27	25.81*

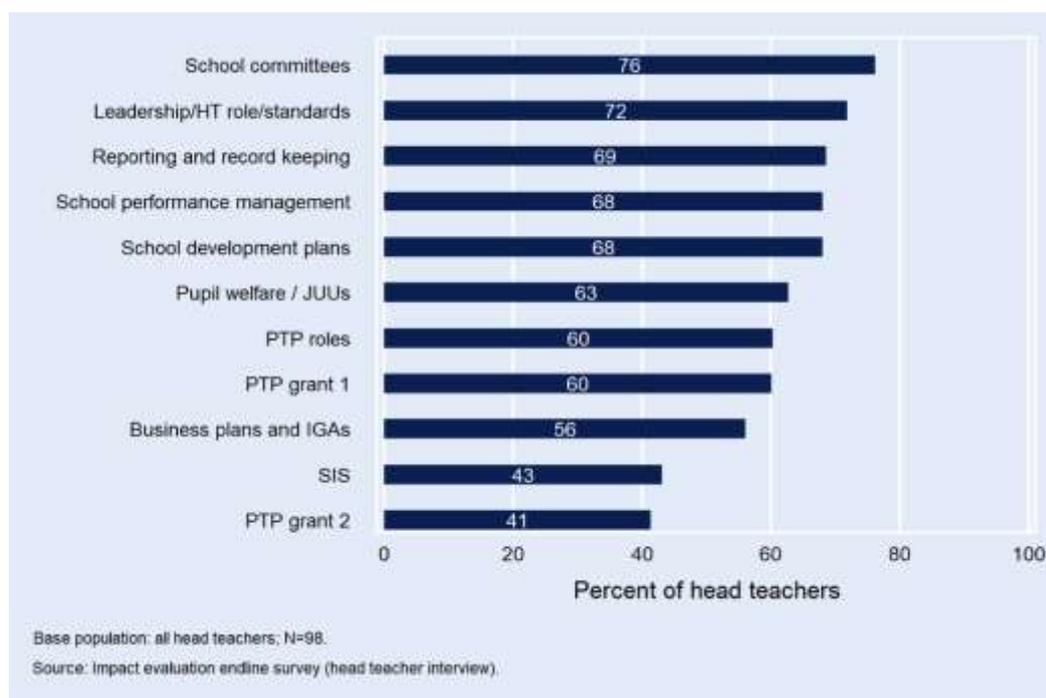
Source: Evaluation midline and endline surveys (head teacher interview).
Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (2) The sample sizes do not sum to 98 (the number of treatment head teachers for whom these data are available) because there is overlap for head teachers who were at the same school at baseline, midline, and endline, and those who were at the same school at midline and endline.

Among head teachers who were at the same school at baseline, midline, and endline as well as those at the same school at midline and endline, only 65–70% received the training on SLM provided **before midline**, whereas close to all (96–98%) received the SLM training provided **after midline** (Table 11).¹⁰² By contrast, among new head teachers at endline, a much smaller proportion (55%) received the SLM training **after midline**, and this is because they were promoted to head teachers after the training had been conducted or because they sent teachers to attend the SLM training on their behalf.

The adverse effects of high head teacher turnover—weakening of the benefits of EQUIP-T SLM training and disruption of SLM in general—were already raised in the EQUIP-T impact evaluation midline report (OPM, 2016b).

The EQUIP-T SLM training has covered the intended topics but a large proportion of head teachers have not received training on all the topics. Figure 15 shows that the two topics most commonly received were SCs' roles and responsibilities (76%), and school leadership roles and school standards (72%); while the two topics least commonly received were SIS (43%); and PTP grant 2 application and management (41%).

That large groups among the current head teachers did not receive training on certain topics may affect the impact of EQUIP-T on expected changes in related SLM practices. This is discussed in subsequent sections, as relevant. For instance, that 32% of head teachers did not receive training on SDPs may have influenced the availability, comprehensiveness, and implementation of SDPs (see Section 5.4).

Figure 15: Contents of EQUIP-T SLM in-service training ever received

Head teachers who attended EQUIP-T SLM training between baseline and midline on average received seven days of training, which is largely consistent with the expected number of training days (see Annex B.5). However, for training attended between midline and endline the average number of training days was only five, which is less than expected. Possible reasons for this are that some head teachers sent teachers to attend some of the SLM training events on their behalf, and that head teachers missed some of the training events.

Nearly all head teachers (97%) who attended EQUIP-T SLM training, both before and after midline, considered it useful. At endline, the most common gains from the training reported by head teachers are: awareness of head teacher responsibilities (71%); teacher management (59%); relationship with teachers (50%); confidence in their role as head teacher (44%); and the relationship with parents / community (44%). The proportions of head teachers reporting the last three gains have increased significantly since midline. At the same time, the proportion of head teachers reporting school development planning as a gain from the training has decreased significantly, from 52% at midline to 24% at endline. The other reported gains at endline include: SCs' responsibilities and roles (29%); academic programme management (24%); financial management (24%); reporting / record-keeping (19%); pupil welfare (12%); and support network (11%).

For head teachers who attended EQUIP-T SLM training, the difficulties reported at midline and endline are largely similar, and are almost completely related to the design, arrangement, and location of the training. At endline, the main reported difficulties are: no / insufficient payment (25%); limited training time (25%) – which is a significant increase, from 3% at midline; envy from colleagues (15%); too much content covered (15%) – which is a significant decrease, from 39% at midline; sessions at/on an inconvenient time / day (14%); not enough training material (10%); and transport difficult / venue too far away (8%). Among the head teachers, 2% reported incurring out-of-pocket expenses in order to attend training. Training venues being far way or difficult to reach, and insufficient payment to cover costs related to attending training, may have influenced the number of training events attended by head teachers.

5.3.2 Has EQUIP-T provided early grade teacher in-service training to head teachers? (EQUIP-T input)

At endline, the vast majority (89%) of head teachers have attended some kind of early grade teacher training, compared to none at baseline.¹⁰³ The largest provider of early grade teacher training is EQUIP-T (89%). Other main providers are STEP (35%) and LANES (26%). All the head teachers who attended early grade teacher training by STEP or LANES also attended EQUIP-T training, meaning they received training provided by two, and in some cases even three, different providers. There is also some overlap in training contents, with all three providers covering general teaching methods and the Standards 1 and 2 curricula.

Head teachers in programme schools are also targeted by EQUIP-T for early grade teacher training to strengthen their academic leadership and teaching support to teachers.¹⁰⁴ That 11% of head teachers in programme schools did not receive the EQUIP-T early grade teacher training may undermine EQUIP-T's impact on head teachers' academic leadership and teaching support.

The contents of the early grade teacher training reported by head teachers who attended such training is largely consistent with the EQUIP-T training modules' contents and timing. The most frequently reported contents are (in parentheses the midline estimates are followed by the endline estimates): Standards 1 to 2 curriculum (70% and 42%), a significant decrease; general teaching methods (62% and 55%); early grade Swahili literacy (60% and 58%); and early grade maths / numeracy (49% and 62%). At endline, 23% of head teachers also report receiving training on gender-responsive pedagogy and 14% report receiving training on the Standards 3 and 4 curricula.¹⁰⁵

5.4 Findings: head teacher capacity and SLM practices

5.4.1 Has head teacher capacity changed? (EQUIP-T output)

SDP availability, contents, and implementation

The availability of SDPs increased significantly between baseline and midline, but then decreased between midline and endline.¹⁰⁶ Despite the importance of having an SDP to guide school management and for school transparency, at endline only 49% of schools have one, compared to 68% at midline, and 36% of schools at baseline (Table 12). The increase in schools with an SDP between baseline and midline coincides with the training on SDPs provided by EQUIP-T (see Section 5.3), and the subsequent decline may be partly related to the timing of the SDP training, combined with the head teacher turnover. This is reflected in the decline in the proportion of head teachers who ever received EQUIP-T SLM training reporting that they gained school development

¹⁰³ This is for all head teachers and among these 11% did not attend any in-service early grade teaching training at endline.

¹⁰⁴ Head teachers are not meant to receive training on the same number of modules as teachers, or for the same number of days.

¹⁰⁵ The latter is a reporting error as the Standards 3 and 4 curricula was covered by LANES training, not by EQUIP-T training. A large minority of head teachers attended both EQUIP-T and LANES training, and both are delivered by government officials, so it is not entirely surprising that there is some recall error related to specific topics.

¹⁰⁶ To collect data on whether schools have an SDP enumerators asked head teachers to see the school plan for 2018. Schools were considered to have an SDP if the school plan shown to the enumerator, and reviewed by the enumerator together with the head teacher, included a number of major activities planned to be conducted during the year and was written down. The plan could be either a draft or final version.

planning skills. This suggests that unless regular refresher training is provided, schools may not continue to prepare SDPs.

Table 12: Availability of SDPs (trends in programme areas)

Indicator	Baseline		Midline		Endline		Difference	
	Estimate	N	Estimate	N	Estimate	N	BL-EL	ML-EL
Has SDP for current school year (% schools)	36.37	100	67.74	100	48.59	100	12.22	-19.15**

Source: Evaluation baseline, midline, and endline surveys (head teacher interview).
Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (2) This is for all schools.

Although the proportion of schools with an SDP has declined since midline, on average SDPs contain more of the core elements at endline. Three elements in particular—teaching and learning objectives; baseline data and targets; and a budget—are considered core features of SDPs. Among the schools with an SDP at endline, 19% contain one of the core elements and a further 19% have two of the core elements, which are significant increases since baseline; and only 2% have all three core elements. No school has an SDP with none of the core elements, which is a significant improvement since both since baseline and midline. This indicates a possible increase in the quality of SDPs at endline for schools that do have them. Still, only 21% of these SDPs contain two or three of the core elements.

At endline, compared to at baseline, a significantly larger proportion of SDPs include: improvements to school facilities (38%); budgets (32%); teaching and learning objectives (28%); strategies to reduce pupil absenteeism and dropout (19%); and strategies to improve girls' learning (9%). Between midline and endline, there were significant declines in the proportions of SDPs including strategies to improve Standards 4 and 7 examination scores (from 15% to 4%), and baseline data and targets (from 9% to less than 2%). 13% of SDPs at endline contain strategies for specific groups of vulnerable pupils and 4% have strategies to improve pupil welfare.¹⁰⁷ That only 2% of SDPs at endline contain baseline data and targets, a decline since midline, is at odds with EQUIP-T's aim of promoting the use of data for school decision making and to improve SLM.

5.4.2 Have head teachers' SLM practices changed? (EQUIP-T intermediate outcome)

This section present findings on head teachers' SLM practices at endline and changes over time. The findings on SPMs, including their frequency and topics discussed during the meetings, were presented in Section 4.2.4 in Chapter 4, and are not discussed here.

SDP implementation

Among all programme schools, 41% have an SDP and have started implementation of it.¹⁰⁸ This suggests that once schools have developed an SDP they use it, and that the challenge is to encourage schools to prepare SDPs to begin with.

¹⁰⁷ The definitions of the terms 'vulnerable children' and 'pupil welfare' used in this study are given in the Box at the start of Chapter 7.

¹⁰⁸ For implementation of the current school year's SDP to be considered to have started, implementation of at least one activity in it needed to have begun. Head teachers were asked to show documentation to support this, in the form of school meeting minutes or financial records, if available.

Teacher management

The only significant change in the main factor used by head teachers to assess teacher performance since baseline has been the increase in the use of pupils' academic results.

This has risen from 18% at baseline to 22% at midline to 39% at endline. The other two main factors at endline are: checking teachers' lesson preparations (30%) and checking teachers' performance in class (19%). The least reported factor is monitoring of teacher attendance and punctuality (3%), which is in line with the observed high levels of teacher absence from school and the classroom, as well as late arrival at school (see Section 4.7.1 in Chapter 4).

Head teacher checking of lesson plans has increased significantly since baseline. At endline, 96% of Standards 1 to 3 teachers report that the head teacher checks their lesson plans, compared to 94% at midline and 91% at baseline (Table 13). Only about 20% of head teachers provide written feedback on teachers' lesson plans to help improve lesson planning at endline, and this has not changed significantly since midline (24%).¹⁰⁹

Observation of lessons by head teachers has decreased significantly since midline. The proportion of Standards 1 to 3 teachers reporting that head teachers observed their lessons during the last 30 days was already low at midline, at 37%, and has declined significantly, to 15% at endline (Table 13).¹¹⁰ However, lesson observations can also be conducted by others, including the assistant head teacher, academic master, senior teachers, and peers, suggesting that some head teachers delegate this responsibility. If lesson observations by others are combined with those by head teachers, 23% of Standards 1 to 3 teachers at endline were observed in the last 30 days compared to 44% at midline, which is a significant decline. This clearly shows that fewer lessons are observed, whether by the head teacher or other school staff, at endline than at midline.

Written feedback from head teachers following lesson observations is rare. At midline 6% and at endline 3% of Standards 1 to 3 teachers reported receiving written feedback from head teachers on their observed lessons.

Table 13: Teacher management practices reported by Standards 1 to 3 teachers (trends in programme areas)

Indicator	Baseline		Midline		Endline		Difference	
	Estimate	N	Estimate	N	Estimate	N	BL-EL	ML-EL
Lesson plans								
Lesson plans checked by head teacher (% Stds. 1–3 teachers)	91.11	327	93.81	384	95.99	418	4.88**	2.18
Written lesson plan feedback from head teacher (% Stds. 1–3 teachers)			23.94	384	20.05	418		-3.89
Lesson observation								
Lesson observation by head teacher (% Stds. 1–3 teachers)			36.96	384	14.66	418		-22.3***
Lesson observation by head teacher or other (% Stds 1-3 tchs)			43.91	384	23.3	419		-20.61***
Written lesson observation feedback from head teacher (% Stds. 1–3 teachers)			5.48	383	2.79	418		-2.69
Performance appraisal								

¹⁰⁹ The question asked was: 'Did you receive any written feedback on your lesson planning from the head teacher or academic master *that includes a substantive comment on the contents of the plan?*' The part in italics was added in the endline survey instrument, whereas at midline this text was not written out in the instrument but the enumerators were very clearly trained on it.

¹¹⁰ The definition used for this question is that head teachers must observe all of the lesson and not just walk in for a few minutes to check that things are going well, and they need to be observing the teacher's teaching behaviour.

Received at least one performance appraisal in previous school year (% Stds. 1–3 teachers) ²	27.52	327	28.33	384	29.68	417	2.16	1.35
---	-------	-----	-------	-----	-------	-----	------	------

Source: Evaluation baseline, midline, and endline surveys (teacher interview).
Notes :.(1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (2) This is for Standards 1–3 teachers. (2) This indicator is also used as an impact indicator; see Volume II Chapter 5 for details.

Head teachers' use of performance appraisals has not changed significantly since baseline.

At baseline, midline, and endline, only around 28–30% of Standards 1 to 3 teachers reported receiving at least one performance appraisal during the previous school year to discuss their performance and professional development needs. The lack of change for this particular practice is not unexpected, because although EQUIP-T developed a Leadership Competency Framework before midline as a basis for performance appraisal (see Annex B.3), implementation stalled, and the approach was later superseded by materials developed together with the Agency for the Development of Education Management (ADEM). These materials were in their pilot phase at the time of the endline survey. Consistent with this, there is no evidence of any impact of EQUIP-T as a whole on the use of teacher performance appraisals (see Volume II, Chapter 5).

Regular staff meetings between the head teacher and teachers have increased significantly since baseline.

Staff meetings are typically chaired by the head teacher and attended by teachers, and sometimes also non-teaching staff, to discuss administrative and other school matters. At baseline, 15% of Standards 1 to 3 teachers reported they attended four or more staff meetings during the last 60 days, by midline this had increased significantly to 32%, but it then declined somewhat, to 27%, at endline. On average, three staff meetings were held over the last 60 days, with no significant change over time.

The existence of rewards for teachers who perform well has become significantly more common since baseline according to head teachers. At endline, 55% of head teachers report that their school has performance rewards for teachers who perform well, compared to 46% at midline and 34% at baseline. The most common types of rewards reported by head teachers are: financial incentives (34%), which is a significant increase since both baseline and midline; verbal recognition (26%); certificate, cup, or medal (8%); and material rewards (6%). A lower share of Standards 1-3 teachers report schools providing rewards (32% at endline).

The large majority of head teachers (80%) report that action is taken if teachers perform poorly, and this has not changed significantly over time. At endline, the most common actions according to head teachers are: a warning is given by the head teacher (62%); the head teacher reports to the WEO or a warning is given by the WEO (15%); a warning is given by the SC (13%); and an increase in lesson observations (5%). Standard 1 to 3 teachers' views on whether schools sanction them for poor performance are broadly similar to head teachers (73% said that action was taken for poor performance at endline).

Most head teachers consider teacher attendance at their school to be 'good' or 'very good' (94% at midline and 92% at endline). Moreover, 56% of head teachers think that teacher attendance at their schools is better today than two years ago; 26% think it is similar; and 18% think that it has worsened. This is in conflict with the measured teacher absence from school, which remained virtually the same between midline (14%) and endline (13%) (see Section 4.7.1 in Chapter 4). At the same time, the majority of head teachers (71%) report that teachers are sometimes absent from the classroom, which is in line with the high teacher classroom absence in the programme schools (see Section 4.7.1 in Chapter 4).

Education improvement actions taken by head teachers

At endline all head teachers report taking action to improve education in the school in the previous school year, similar to at midline. Taking action to increase pupil attendance in the previous school year has risen significantly from midline to endline, 17% to 45%, although levels of pupil attendance did not improve over the same period.¹¹¹ At the same time, taking action related to TLMs has declined significantly, from 21% to less than 1%. One contributing explanation may be that a significantly larger proportion of schools received textbooks during the school years 2016 and 2017 (100%) than during the school years 2014 and 2015 (74%) (see Section 5.6.1). Comparing this to reporting by Standard 1 to 3 teachers, 81% of teachers at endline, and 75% of teachers at midline, said that head teachers took action to improve education the previous school year, which is lower but still indicates that a large majority of head teachers took some action.

The other main actions reported by head teachers are: providing extra tuition classes (38%); strengthening relationships with parents and the community (34%); giving more tests / exams (26%); improving school infrastructure (20%); improving teacher attendance and punctuality (15%); extra-curricular activities (9%); and school feeding (9%). Only a very small group of head teachers report taking actions to fundraise (3%); conducting IGAs (2%); and improving pupil welfare (1%).

5.5 Findings: head teacher absence and job satisfaction

5.5.1 Head teacher absence and job satisfaction (Equip-t output to intermediate outcome assumptions)

For increased head teacher SLM capacity to translate into improved SLM practices, head teachers must be present at school, but head teacher absenteeism is high. At endline, 21% of head teachers were absent on the day of the survey using a head count, and this does not represent a significant change since either baseline (16%) or midline (15%). This may undermine EQUIP-T's impact on SLM, regardless of the reasons for head teachers being absent.

The large majority of head teachers (94%) report that they were absent from school during the last 30 days, which is a significant but relatively small decrease since midline (100%).¹¹² Official and work-related reasons for head teacher absence are by far the most commonly reported: attending meetings (76%) and official Government work not related to education (29%). There has been a significant and large reduction in head teachers reporting attending training as a reason for absence, from 33% at midline to 2% at endline. This is likely partly related to the new head teachers at endline who joined their schools after the completion of the EQUIP-T SLM training (see Section 5.3). The most common personal reasons for absence are: illness (17%), which has increased significantly since midline, and family responsibility (14%).

A large proportion of head teachers (63%) report having outstanding non-salary claims – for example, for leave, transfer, or sickness. This may affect head teacher motivation and thereby attendance.

Head teachers were asked at baseline, midline, and endline on the day of the survey how satisfied they feel with their job as a head teacher. The average rating for job satisfaction is eight out of 10, where one means 'completely unsatisfied' and 10 means 'completely satisfied', and this has remained the same over time. When asked if their job satisfaction has changed compared to two

¹¹¹ See Chapter 3, Section 3.2.2 Table 3 for trends in early grade pupil absenteeism.

¹¹² This is head teacher self-reported absence from school, and over the last 30 days, which is not comparable to the head teacher school absenteeism estimate based on the head count.

years ago, 50% say it is higher; 24% that it is similar; and 26% that it is lower; this has not changed since midline. Based on these findings, on average, head teachers seem relatively satisfied with their jobs. However, at midline the qualitative research that went into more depth on this issue found that head teachers felt unmotivated and that their morale was low (OPM, 2016b).

5.6 Findings: capitation grant payments and in-kind resources

5.6.1 Capitation grant payments received in full and in-kind resources received (EQUIP-T output to intermediate outcome assumption)

Prior to the school year 2016, capitation grant payments were transferred from the Government via district authorities to schools, and payments were erratic and typically less than the stipulated amounts. In January 2016 the payment mechanism was changed so that capitation grants flow directly from the central Ministry of Finance to schools' bank accounts on a monthly basis. The expected capitation grant is Tanzania shilling (TZS) 6,000 per pupil per year.¹¹³ Based on the data collected at midline and endline, the average received capitation grant amount has increased significantly, from TZS 1,200 per pupil in 2015 (which is a large shortfall compared to the expected amount) to TZS 6,500 per pupil in 2017 (which is slightly higher than the expected amount).

In 2015 none of the programme schools received the capitation grant in full, whereas in 2017 61% did. This significant and large improvement means that most schools now have access to more reliable funding, and to the expected amount, which is a key assumption in the programme theory of change as regards facilitating better school development planning and implementation.

There has been a significant increase in the proportion of schools receiving in-kind resources since baseline. At endline all schools have received in-kind resources during the previous two school years, compared to 88% at baseline. There has been a significant and massive increase in schools receiving classroom furniture, from 24% at baseline to 31% at midline to 81% at endline, in line with the Government directive to increase the availability of desks in classrooms. Classrooms have been constructed for 13% of schools at endline, compared to 8% at baseline and 2% at midline, and 10% of schools have received electricity, up from 2% at midline. Receipt of TLMs by schools is discussed in Section 4.6 in Chapter 4.

School characteristics and infrastructure influence SLM. For example, large schools with very large class sizes and classroom shortages place different demands on head teachers than smaller schools with appropriate class sizes and adequate infrastructure. Box 12 discusses the challenging context faced by head teachers in many of the programme schools.

Box 12: Key school characteristics and infrastructure

The **average school size** at endline is 597 primary pupils per school, and has risen significantly from baseline (486) and midline (472). This increase is almost certainly related to the introduction of fee-free primary education at the end of 2015.

A shortage of classrooms is a major challenge that was already an issue at baseline and that has become even more so at endline. On average, for Standards 1 to 7 there are 86 **pupils for every classroom in use**, which is a significant increase since both baseline and midline, and much higher than the national benchmark of 45 pupils per classroom (MOEVT, 2009). There are also considerable differences between schools: the 10% of schools with the smallest classroom shortage have 48 pupils per classroom (still above the benchmark) while the 10% of schools with the largest classroom shortage have 136 pupils per

¹¹³ Equivalent to approximately £2.00 and US\$2.60 at current exchange rates.

classroom, and 75% of schools have 60 pupils or more per classroom in use. This clearly shows that classroom shortages are acute in many programme schools.

Schools can address classroom shortages in two main ways: by having two shifts of classes using the same classroom at different times (double-shifting), or by teaching multiple classes in the same classroom at the same time. The Primary Teacher Deployment Strategy, to be implemented from 2018, includes a policy of operating second shifts for Standards 1 to 4 for primary education (MOEST, 2018). At baseline 48% of schools used **double-shifting of classes**, by midline this had increased significantly to 66%, and it has increased even further by endline, to 73%, which likely reflects both the practical necessity of using second shifts and the new policy. Double-shifting tends to reduce instructional hours, which in turn affects pupil learning.

New EQUIP-T activities related to school infrastructure were introduced in 2017 as part of the programme extension, including community-led construction of up to 220 satellite schools and 230 finished classrooms including roofing (EQUIP-T MA, 2017). These activities are intended to help address the classroom shortages.

Table 14: Selected school characteristics and infrastructure (trends in programme areas)

Indicator	Baseline		Midline		Endline		Difference	
	Estimate	N	Estimate	N	Estimate	N	BL-EL	ML-EL
Total pupil enrolment (mean)	486.42	100	472.35	100	597.11	100	110.69***	124.76***
Number of pupils per classroom in use (mean)	75	92	72.43	92	86.1	92	11.1***	13.67**
Schools with more than 60 pupils per classroom in use (%)	58.61	92	56.51	92	75.16	92	16.55*	18.66**
Second shift	47.89	100	66.47	100	73.33	100	25.44**	6.86
Number of pupils per functional toilet (mean)	74.08	94	69.86	94	93.19	94	19.12**	23.33**
Drinking water on school premises (% schools)	31.93	100	34.8	100	31.92	100	-0.01	-2.89

Source: Evaluation baseline, midline, and endline surveys (head teacher interview and school records checks).
Notes: (1) Asterisks indicate statistical significance levels: *** p<0.01, ** p<0.05, * p<0.1. (2) This is for all schools.

School sanitation has deteriorated since baseline, as the average number of **pupils per functional toilet** has risen from 74 at baseline to 93 at endline. Only 32% of schools have **drinking water** available on school premises and this share remains similar to the value at baseline and midline. On the positive side, 21% of schools have **functional electricity** at endline, compared to only 5% at baseline, although this is still a very small proportion.

5.7 Findings: SIS training and use

5.7.1 Has EQUIP-T trained head teachers on SIS and provided SIS tablets? (EQUIP-T input)

All head teachers were meant to receive training on SIS management after midline but only 43% of current head teachers did. The reason for such a large proportion of current head teachers not receiving SIS training may be the high head teacher turnover (see Box 11). Irrespective of the cause, this will weaken head teachers' ability to use the SIS, and will undermine any EQUIP-T impact on SLM coming through the new SIS.

EQUIP-T is the only provider of SIS tablets in the programme schools, and 97% of schools have received a tablet during the last two years. At endline, 94% of schools report that they have a functioning tablet, and the tablet was seen by enumerators in 70% of the schools, but tablets were only charged in 55% of schools, despite solar chargers being supplied with the

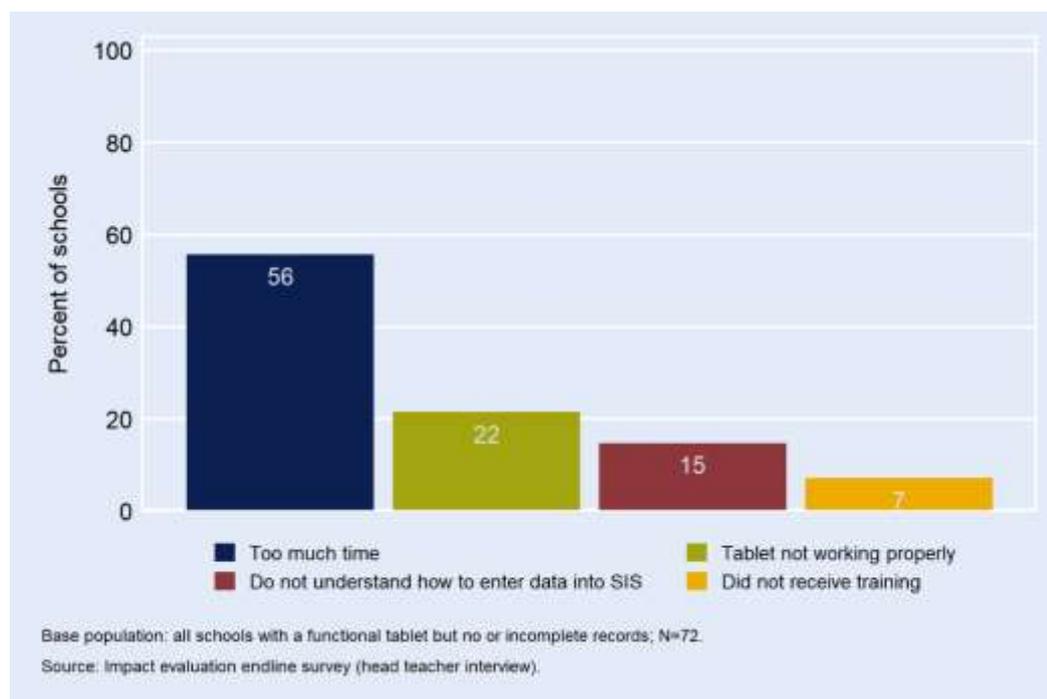
tablets. That only 21% of schools have functional electricity (see Box 12) means that it may not be straightforward to charge tablets, and charging tablets may be a slow process, which creates a barrier to head teachers entering and regularly updating pupil, teacher, and other school data in the SIS.

5.7.2 Is the EQUIP-T SIS functional? (EQUIP-T output)

Only 24% of schools with functional tablets have entered up-to-date lists and data for both teachers and pupils in the school at the time of the endline. For these schools it took on average 17 hours to enter these data.¹¹⁴

According to head teachers in schools that have functional tablets but no or incomplete records, the main reasons for having no or incomplete records are as illustrated in Figure 16. The time taken to enter the data is the top reason for non-compliance, and this may reflect issues with the design of the application. That tablets are not working properly and head teachers not understanding how to enter data into the SIS and / or not receiving training, are factors that it is within EQUIP-T's scope to address.

Figure 16: Reasons for schools not entering data on pupil enrolment and current teachers in the SIS



Very few schools are using the tablets to enter pupil and teacher attendance data into the SIS. Only 10% of schools with functional tablets have entered pupil attendance data for all pupils in all classes into the tablet **on at least one day** during the current school year, while this is expected to be a daily tool. The main reasons for these schools not entering pupil attendance data are that it takes too much time (13%) and that the tablet was not working properly (7%).¹¹⁵ For teacher

¹¹⁴ This does not include entering attendance data but is only about how long it took to add the lists of all the pupils and teachers at the school into the tablet and record information about them. The entered data were checked by the enumerators. 94 out of the 100 schools had a functional tablet.

¹¹⁵ This is for schools with functioning tablets so this refers to other issues than the tablet itself being broken.

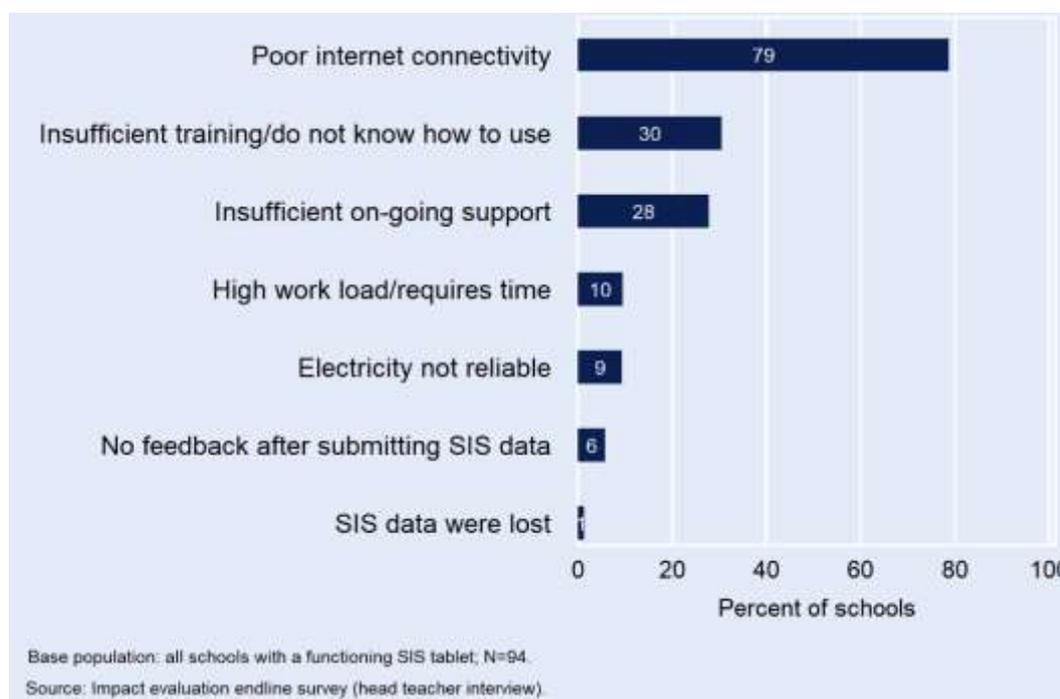
attendance, merely 17% of schools with functional tablets had entered such data into the tablet for all teachers on at least one day during the current school year.

5.7.3 EQUIP-T SIS tablet software is fit for purpose and tablet-based SIS is appropriate (EQUIP-T input to output assumption)

For schools that have functional SIS tablets, the reported difficulties with using them (Figure 17) that it is hard for EQUIP-T to influence are: poor internet connectivity (79%);¹¹⁶ high workload and time it takes to enter the SIS data (10%, which may partly be a design issue indicating the need to simplify the SIS application and / or reduce the scope of the data to be entered); and regular electricity supply to charge the tablet (9%). The latter suggests that there may have been some problems with the supply of solar chargers to accompany the tablets.

Reported difficulties that EQUIP-T could directly influence are: insufficient direct training on the SIS and head teachers not understanding how to use the SIS / tablet (30%); insufficient ongoing support (28%); and schools not receiving feedback after submitting SIS data (6%).

Figure 17: Difficulties in using SIS tablets



5.7.4 Does the SIS provide useful reports to support SLM? (EQUIP-T intermediate outcome)

At this stage the SIS is not providing useful reports to support SLM, largely because a majority of schools have entered incomplete data or no data at all. Only 9% of schools with functional tablets use the SIS for SLM purposes or community engagement, and 9% use it for discussions with WEOs. The SIS has replaced other records and reports in 24% of these schools.

¹¹⁶ Data can be entered into the tablets without internet access but connectivity is required to transmit the data to the local SIS database.

Despite most head teachers not using the tablets or SIS to support SLM, 70% find them very useful and 27% find them somewhat useful, possibly for reasons other than entering data into the SIS.

The EQUIP-T MA provided an update on the situation with the SIS initiative, as part of written feedback on the draft version of this report.¹¹⁷ Since the endline survey was completed in May 2018 the programme, PO-RALG and the SIS application developer have worked to develop a more stable version of the application and this was finalised in October 2018. In addition, earlier versions of the SIS application could only generate a limited number of reports using the data entered into the SIS that may not have been considered very useful by head teachers for SLM purposes. Since then, the SIS application has been updated to be able to generate a larger number of reports. Together these adjustments may help alleviate some of the problems with data entry and use of the SIS identified by the survey.

5.8 Findings: WEO support for schools and head teachers

5.8.1 Is WEO support for head teachers effective? (EQUIP-T output, component 3B)

The findings on school quality assessors (SQAs), previously known as district school inspectors (DSIs), and WEOs presented in this section relate to EQUIP-T component 3B, district planning and management, but as they involve external support to schools, including to strengthen SLM, it is useful to discuss them here.

Only about half of schools (52%) have received a visit from an SQA in the previous school year at endline, which is not significantly different from midline or baseline. On average, schools that have received an SQA visit received 1.5 visits. The midline qualitative research found that SQAs visited schools much less frequently than WEOs, and that they were not able to cover all schools every year (OPM, 2016b). By contrast, nearly all schools (98%) received at least one WEO visit in the school year preceding endline, and this was also the case at midline (99%) and baseline (99%).

The number of WEO visits has increased significantly for the majority of schools since baseline. Schools are meant to receive one or more WEO visits per month during each term, for a total of at least 12 visits per school year (EQUIP-T MA, 2016). At endline, 69% of schools have received 12 or more WEO visits in the previous school year, compared to 36% at midline and 9% at baseline, a very large and significant increase in WEO visits over a short period of time.

At the same time, the average duration of WEO visits has declined somewhat since midline, but this change is not significant. At endline 29% of WEOs spent less than half an hour at the school during their last visit; 27% spent between half an hour to one hour; 23% spent one to two hours; and 21% spent more than two hours. During visits that are shorter than an hour it is unlikely that WEOs can provide substantial support to head teachers and teachers.

5.8.2 Have WEO school support practices changed? (EQUIP-T intermediate outcome)

¹¹⁷ The evaluation team received written comments on the draft endline quantitative report from the EQUIP-T MA on 15 November 2018.

WEOs engage in a variety of activities during their visits to schools. For the last school visit the most common activities were: checking school records (25%); bringing exam papers or supervising exams (23%); checking teacher records (21%); observing lessons (17%); observing school management practices (14%); checking pupil work (9%); and coaching or participating in sports (7%). The only significant change since midline is that the proportion of head teachers reporting that WEOs brought exam papers with them or supervised exams more than doubled from 10% to 24%.

During their last visit WEOs provided advice or support in 61% of schools at endline. The most common areas for WEOs to provide advice on were: teaching and learning (26%); pupil attendance and punctuality (16%); teacher attendance and punctuality (11%); the SC (7%); and community and parental engagement (7%). Advice on pupil welfare and inclusive education for girls, respectively, was provided in less than 2% of schools.

There has been a significant and large decline in the proportion of head teachers who consider the WEO's last visit to have been 'very helpful', from 76% at midline to 56% at endline. Moreover, at endline, 80% of head teachers consider the WEO support to their school to be 'good' or 'very good', which is a decrease from 91% at midline. The reason for this change in head teachers' views of WEO may be related to the very high WEO turnover since 2016 (see Section 5.8.3 below).

5.8.3 WEO turnover (EQUIP-T input to output assumption)

Since the start of 2016, 56% of schools have had a new WEO, according to head teacher reporting, which is a very large change over a short period of time. This is likely the result of the professionalisation drive for WEOs implemented in 2017, under which a higher qualification requirement for WEOs, to have a degree, was introduced. This led to a large number of WEOs having to leave their positions (OPM, 2018). The very high turnover of WEOs is likely, at least in the short run, to affect their effectiveness in supporting head teachers and schools as they need to get used to their new roles, and new relationships have to be built with head teachers and other education staff.

5.8.4 Are head teachers reporting to WEOs / districts? (EQUIP-T output)

The frequency of schools reporting to WEOs and districts on school performance, facilities, and finances has decreased significantly since midline. At midline all schools prepared monthly written reports to the WEO or district. By contrast, at endline only 86% of schools provide monthly reports; 3% provide quarterly reports; and 7% provide annual reports; while 3% do not provide any report at all. This may be linked to the finding that the SIS has replaced reports in some schools.

The contents of school reports to WEOs or districts vary, and the reports on average cover fewer topics at endline than at midline. The most common contents at endline are: teacher attendance (62%); school infrastructure and furniture (62%); pupil enrolment (60%); pupil attendance (57%); other teacher information (52%); school budget or finance information (45%); and school committee information (25%). But at endline significantly fewer schools report on each of these topics, except for pupil enrolment and school committees, than at midline.

5.8.5 Are head teachers attending ward education and COL meetings? (EQUIP-T output)

Ward education meetings are chaired by the WEO and are attended by head teachers from schools in the same ward. They are supposed to happen on a monthly basis and they take place at one of the schools in the ward. These meetings were part of the education system before the start of EQUIP-T, and are being strengthened by the programme.

The proportion of head teachers attending ward education meetings has declined significantly since midline. At midline, 91% of head teachers had attended a ward education meeting during the last 60 days but at endline only 69% had attended such a meeting. On average, head teachers attended three ward education meetings in 2017.

COL for head teachers are peer support meetings, initiated by EQUIP-T, that are intended to be chaired by a head teacher from the ward and attended by other head teachers in the same ward to share good practices and experiences, and to support each other. Head teachers are supposed to initiate these COL and the WEO is not a formal part of the group, although the WEO may attend some meetings. By the time of the endline survey, district-level training on formation of head teachers' COL had taken place, but training of head teachers themselves had not yet happened formally. According to the EQUIP-T MA the SLM COL for head teachers were to start operating in July 2018.¹¹⁸

At the time of the endline survey that was completed in May 2018, 33% of head teachers reported they had attended a head teacher led support meeting with other head teachers during the last 60 days.¹¹⁹ This means that either head teachers are referring to other peer meetings for head teachers that are similar to the intended COL meetings, or that COL meetings started earlier than planned in some districts.

5.9 Summary of findings on SLM

This section summarises the evaluation evidence on the effectiveness of EQUIP-T in achieving its intended SLM inputs, outputs, and intermediate outcomes, and factors that may have influenced their achievement and that may also affect programme sustainability. Table 15 shows to what extent key EQUIP-T inputs, outputs, and intermediate outcomes have been achieved and underlying theory of change assumptions have been met.

Table 15: Summary of evidence on SLM and SIS

Evaluation questions and assumptions related to the theory of change	Result at EL	Significant change	
		BL to EL	ML to EL
Has SLM strengthened in treatment schools, with changes consistent with the theory of change?			
Inputs Has EQUIP-T provided SLM in-service training and SIS tablets?			
Attended EQUIP-T SLM training <i>ever</i> (% head teachers)	80%
Received EQUIP-T SLM training on SDPs (% head teachers)	68%
Received EQUIP-T SLM training on SIS (% head teachers)	43%
Functioning EQUIP-T SIS tablet available (% schools)	94%

¹¹⁸ The EQUIP-T MA shared this information in their feedback on the draft of this report, received 15 November 2018.

¹¹⁹ Head teachers were asked 'Did you attend any other peer support meeting in your ward in the last 60 days?' This is different than the ward education meeting which is chaired by the WEO. The peer support meetings are typically chaired by a head teacher from the ward and attended by other head teachers in the ward to share good practice and experiences and support each other.

Outputs Has head teacher SLM capacity increased?			
Has SDP for current school year (% schools)	49%	none	negative
Pupil attendance entered into tablet on at least one day in current school year (% schools with functional tablet)	10%
Teacher attendance entered into tablet on at least one day in current school year (% schools with functional tablet)	17%
SIS tablet has replaced other records / reports (% schools with functional tablet)	24%
Visited by WEO in previous school year (% schools)	98%	none	none
Received 12 or more WEO visits in previous school year (% schools)	69%	positive	positive
Attended head teacher led peer support meeting in last 60 days (% head teachers)	33%
Theory of change assumptions What factors may have influenced delivery of outputs?			
Head teacher at same school since baseline in 2014 (%)	26%
Poor internet connectivity does <u>not</u> affect use of SIS tablet (% schools with a functional tablet)	21%
Takes too much time to enter data into SIS is <u>not</u> a main reason for having no or incomplete data (% schools with a functional tablet, and no or incomplete data)	44%
WEO is the same as at the start of 2016 (% schools)	44%
Intermediate outcomes Have head teacher SLM practices improved?			
Has SDP for current school year and has started implementation (% schools)	41%	-	-
Lesson plans checked by head teacher (% Stds. 1–3 teachers)	96%	positive	none
Lesson observation conducted by head teacher (% Stds. 1-3 teachers)	15%	-	negative
Received at least one performance appraisal in previous school year (% Stds. 1–3 teachers)	30%	none	none
Head teacher took action to improve education in previous school year (% Stds. 1–3 teachers)	80%	..	none
Used SIS for SLM / community engagement (% schools with functional tablet)	9%
Used SIS for discussions with WEO (% schools with functional tablet)	9%
Theory of change assumptions What factors may have influenced delivery of intermediate outcomes?			
Present at school on the day of the survey (% head teachers, head count)	79%	none	none
Capitation grant received in full, year prior to the survey round (% schools)	61%	-	positive
Notes: (1) .. means not applicable either because it is an EQUIP-T input or because of the nature of the indicator. (2) - means trend data are not available. (3) SPMM findings are presented and summarised in Section 4.2.4 in Chapter 4.			

A relatively large group of current head teachers have not attended any EQUIP-T SLM training during the last four years, although all head teachers in programme schools were expected to. Of particular note is that even larger groups of current head teachers have not received training on SDPs or SIS management. These training shortfalls will weaken any effect of EQUIP-T on head teacher capacity and SLM practices expected to be achieved through the training. The main reason head teachers have not been trained is the high turnover of head teachers in the programme schools, which is beyond the direct control of EQUIP-T but is an enduring feature of the Tanzanian education system.

The development and use of SDPs to guide school management and increase transparency is a key output but less than half of schools have one at endline. There was a significant increase in schools with an SDP between baseline and midline that coincided with the EQUIP-T training on SDPs, followed by a significant decline since midline, which is likely related to the timing of the SDP training, combined with the high head teacher turnover. This raises concerns about the sustainability of SDPs as an SLM tool after the programme ends.

There has not been any significant overall improvement in head teachers' SLM practices since baseline. Head teachers' use of performance appraisals has not changed significantly over

time, and remains low. When it comes to head teachers observing teachers' lessons, this practice has declined significantly since midline, and at endline only a small group of teachers report that head teachers observe their lessons. At endline all head teachers report taking action to improve education in the school, similar to at midline. The one area where SLM has improved somewhat is the fact that close to all head teachers check lesson plans at endline, which is similar to the situation at midline, and is a significant but small increase since baseline.

EQUIP-T has delivered SIS tablets to close to all programme schools during the last two years, and in the large majority of schools these tablets are functional at endline. However, given that many head teachers have not received training on SIS management, any EQUIP-T impact on SLM practices through the use of the new SIS will be weakened.

The programme schools are not entering pupil and teacher attendance into the SIS anywhere near as frequently as intended, if at all. The main reported difficulty in using the SIS tablets is by far poor internet connectivity, followed by insufficient direct training on the SIS and insufficient ongoing support. The time taken to enter data is another obstacle. Unless these issues are addressed, any impact of EQUIP-T that is expected to come through head teachers using the new SIS to support SLM will be limited. This also raises the fundamental question of whether a tablet-based SIS that relies on good internet connectivity to transfer data is a feasible approach in this context.

At this stage the SIS is not providing useful reports to support SLM, mainly because the large majority of schools have entered incomplete data or no data at all. In schools with functional tablets, these have replaced other records and reports in less than one-quarter of schools. However, only a very small group of schools use the SIS for SLM purposes and community engagement or discussions with WEOs.

One factor contributing to the lack of change in SLM practices, including the use of the SIS, is undoubtedly the high head teacher turnover, as it disrupts the link between head teachers receiving training, increased capacity, and improved SLM practices. Other factors that may prevent increased head teacher capacity translating into improved SLM practices is head teacher absenteeism from school, which is high and has not changed significantly since baseline or midline. The lack of improvement in overall SLM is also despite the majority of schools receiving capitation grant payments in full in 2017, compared to none in 2015, which would be expected to facilitate school development planning and actions to improve education quality.

External support for head teachers and schools through WEOs has increased significantly and by a large amount since baseline. Schools are meant to receive at least 12 WEO visits per school year, and at endline the large majority of schools have, compared to only a small group of schools at baseline. This positive trend was visible already at midline and has continued at endline. Regular and frequent WEO visits should help enhance the sustainability of external school support and training gains after the end of EQUIP-T, assuming that WEOs have the capacity and means to provide effective support.

Although WEOs engage in a variety of activities and provide advice and support, significantly fewer head teachers find the WEO visits very helpful at endline compared to at midline, and they rate the support received lower. This may be related to the high WEO turnover since the start of 2016, related to the WEO professionalisation drive in 2017, which may in the short run be affecting WEOs' effectiveness in providing support to head teachers. WEOs' ongoing support is an essential part of EQUIP-T's sustainability strategy, and therefore it is crucial to ensure WEOs are motivated and have the required skills to effectively support head teachers and schools.

6 Community participation and accountability

Under EQUIP-T component 4A, the activities aim to increase and strengthen community participation and accountability in education, and to promote the sustainability of achieved changes after the programme finishes (EQUIP-T MA, 2017). This first part of the chapter sets out the results chain for community participation and accountability from the programme theory of change, implementation progress since baseline, and expected changes as a result of the activities. Next, the quantitative findings are presented by theme and level in the theory of change (note that schools' accountability to parents and the wider community is not covered by the quantitative research). To finish, the main community findings are summarised, and an overview of the degree to which different inputs have succeeded in achieving the planned outputs and intermediate outcomes is provided.

Box 13 describes some of the key community concepts and partners for EQUIP-T.

Box 13: Description of key community concepts and partners

School committee (SC): Formal school governance body that has a leadership role in school management. It consists of community members and school staff, including teachers. Typically, the head teacher is the secretary and a community member is the chairperson. Its main responsibilities are to: address school needs and solve problems both within the school and between the school and the community; approve the school budget; facilitate IGAs; develop and implement SDPs; ensure pupils' attendance and high academic performance through monitoring; educate parents on the importance of education; and monitor teacher attendance and commitment.

PTP: Class-based groups of parents and teachers established to bring parents closer to the classroom. PTPs are intended to consist of one father and one mother from each of the seven standards, and seven teachers. The focus is on classrooms, with parents actively supporting classroom activities and helping to address problems such as pupil absence and dropout, and teacher absence. PTPs are also meant to come up with their own priorities based on specific class needs.

IGA: An activity to generate income or resources in-kind for the school in addition to the funds it receives from standard sources, such as capitation grants. For instance, a school may set up an agricultural business to plant, harvest, and sell vegetables, to generate additional income for the school.

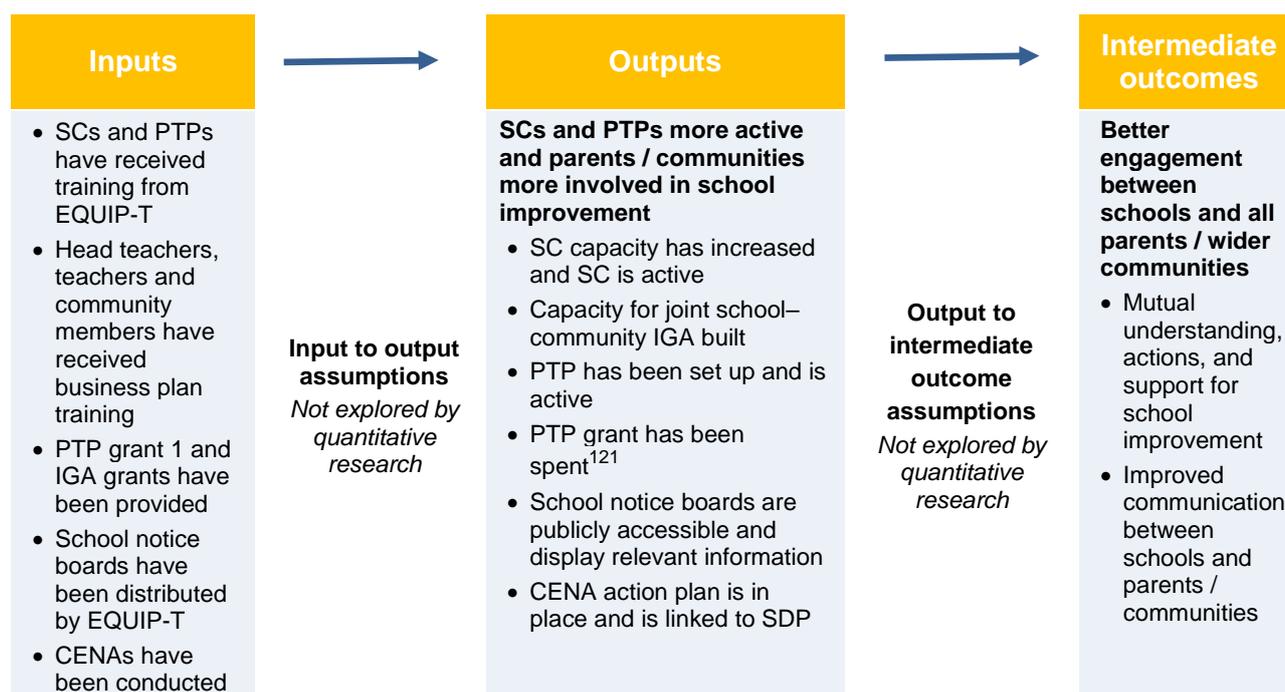
Community education needs assessment (CENA): Assessment of the school's needs conducted by community facilitators trained by local civil society organisations (CSOs), sponsored by EQUIP-T and community leaders, independent of head teachers and SCs. The aim is to develop community action plans that feed into SDPs.

6.1 EQUIP-T component 4A theory of change, implementation, and expectations of change

6.1.1 Theory of change

Increased and strengthened engagement of communities in education is seen as critical for school improvement, and for programme sustainability (EQUIP-T MA, 2017). The main results chain for EQUIP-T component 4A, community participation and accountability, is shown in Figure 18. This excludes theory of change assumptions for this component,¹²⁰ as these are not covered by the quantitative research.

¹²⁰ The theory of change assumptions for this component can be found in the Endline Evaluation Matrix, Part I (Annex C)

Figure 18: Main results chain for EQUIP-T component 4A community

6.1.2 Implementation since baseline

The core activities carried out between baseline and midline were: training for SCs on roles and responsibilities, and school improvement; training on PTP formation and grant management, with the intention that SCs would then oversee the setup of PTPs; provision of PTP grant 1 of TZS 550,000,¹²² of which TZS 100,000¹²³ was for PTP activities and the remainder for general school improvement; distribution of school notice boards and support materials; and training of community facilitators by CSOs to support the development of CENAs and action plans (OPM, 2016a).

Since midline, the main activities have been: training for SCs and PTPs on the roles of the SC, versus the PTP, and on policy, laws, resource management, school mission, and vision, in collaboration with ADEM; refresher training for PTPs; training on business plan development for head teachers / teachers and community leaders; provision of IGA grants of TZS 1,500,000¹²⁴ to 50% of schools in each programme LGA; provision of PTP grant 2 for girls' education activities of TZS 550,000¹²⁵ (discussed in Chapter 7 related to component 4B); and continuation of CENAs and action plans in some communities (EQUIP-T MA, 2017; OPM, 2018).¹²⁶ At the time of the endline survey, implementation of the community score cards with school data from the new SIS meant to enable communities to hold schools to account and identify areas of improvement had not yet begun.¹²⁷

6.1.3 Expected changes

¹²¹ To spend PTP grants, schools had to follow standard Government financial procedures.

¹²² Equivalent to £187 or US\$240 at current exchange rates.

¹²³ Equivalent to £34 or US\$44 at current exchange rates.

¹²⁴ Equivalent to £510 or US\$655 at current exchange rates.

¹²⁵ The EQUIP-T MA clarified that children with disabilities and other marginalised children could also benefit.

¹²⁶ The EQUIP-T activity to develop community score cards and piloting are not covered by this report.

¹²⁷ This is according to the monthly EQUIP-T monitoring reports from regions for January to June 2018.

There are two main changes expected by EQUIP-T staff from the activities implemented since baseline under this component. First, that SC members have become more active and that they participate in and support school activities. Second, that PTPs exist, are active, and support teachers, general school activities, and development, including improvements to the school environment and infrastructure (OPM, 2018).

Some of the EQUIP-T logframe indicators also seek to capture expected changes under component 4A, in particular parents' engagement in school improvement through active SCs and PTPs. The logframe also has an indicator for the community score cards intervention but this has not been implemented at school level yet. For more on the EQUIP-T logframe indicators see Annex B.6.

6.2 Findings: SCs

The findings sections in this chapter present and discuss quantitative evidence on changes in community participation in education and school and community communication since baseline, using data from the evaluation's three survey rounds.

6.2.1 Has EQUIP-T provided training for SCs? (EQUIP-T input)

All schools have a SC at endline, similar to the situation at midline. EQUIP-T has collaborated with ADEM since midline, with the aim of training all SCs in the programme regions (EQUIP-T MA, 2017).

The proportion of SCs that have received some training in the previous two years has increased significantly since midline but training coverage is still not complete. Some 72% of SCs received training **before midline**, provided by EQUIP-T, WEOs, head teachers, or other Government officials, on SC responsibilities and roles, school improvement, and PTP establishment and grants (see Section 6.1.2). **Between midline and endline**, 87% of SCs received training on the roles of SCs versus PTPs; and on policy, laws, resource management, and school mission and vision. These shortfalls in training coverage, in particular for the training provided before midline (22% of SCs not trained), are likely to reduce any effect of EQUIP-T on school and education quality that is meant to come through SCs.

6.2.2 Has SC capacity increased? (EQUIP-T output)

At endline, 59% of head teachers consider the contributions and support to the school provided by the SC to be 'good' or 'very good'. This is not significantly different to the situation at midline (76%) and baseline (54%), although it is a large change.¹²⁸

Some 69% of SCs met during the last quarter at endline, compared to 76% at midline, which is not a significant change.¹²⁹ At midline, several SCs indicate they find it difficult to meet regularly because of a lack of incentives for SC members to attend meetings (OPM, 2016b). When SCs do meet, the vast majority take minutes during their meetings at endline (91%), which is similar to midline (84%) and baseline (91%).

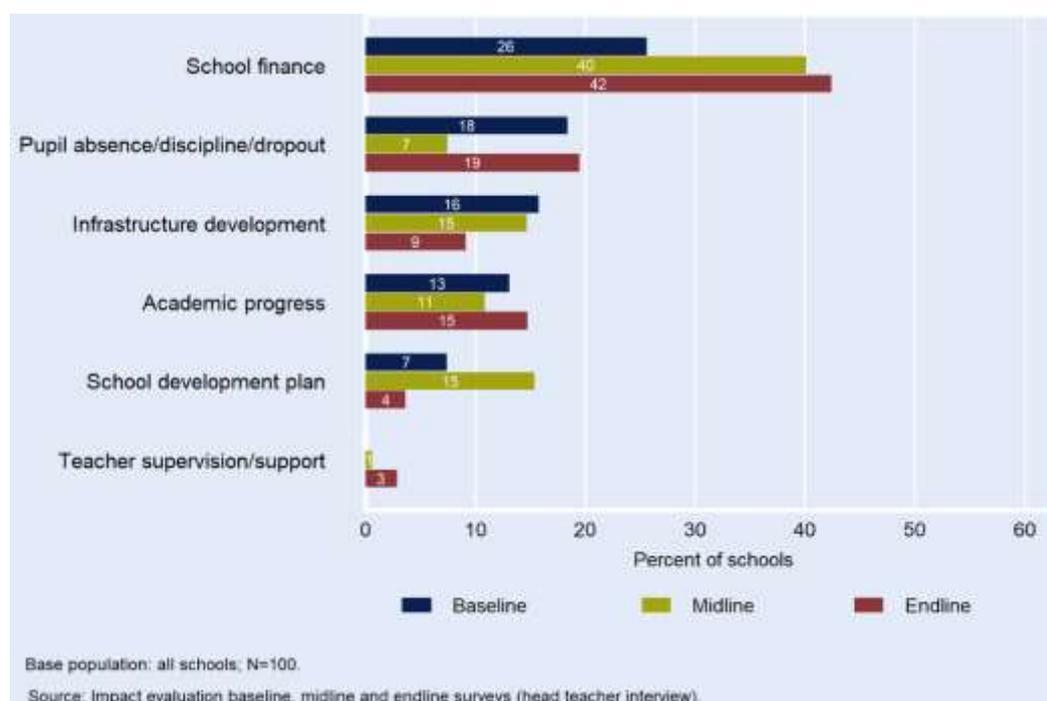
¹²⁸ Head teachers are part of SCs (in the role of secretary), so their views on the support that the SC provides to the school may be influenced by this.

¹²⁹ SCs are meant to meet at least quarterly.

The main topics discussed in SC meetings give an indication of SCs' priorities (Figure 19). At endline the main topic during the most recent meeting was school finances, including parental contributions. This was discussed in 42% of meetings, which is a significant increase from 26% at baseline but is similar to midline. This change is likely related to capitation grants now being received more regularly and in the expected amounts by the majority of schools, and the role of SCs in approving school budgets. This is followed by discussion of pupil absenteeism, discipline and / or dropout (19%), which has increased significantly from midline (7%) but is similar to baseline (18%); and discussion of academic progress (15%) and infrastructure development (9%), which has not changed significantly since midline or baseline.

Another important factor for education quality is teachers but teacher supervision and support was the main topic in a mere 3% of recent meetings, while teacher discipline was not the main topic in any recent meeting. This may be because these are sensitive topics that are affected by the relationship between parents and teachers. Overall, it appears that school finances and infrastructure, on the one hand, and pupil learning, academic progress, and attendance, on the other, are two key concerns for SCs.

Figure 19: Main topic discussed during SC meetings (trends in programme areas)



6.3 Findings: PTPs

6.3.1 Has EQUIP-T provided training for PTPs? (EQUIP-T input)

EQUIP-T's purpose in supporting the establishment of PTPs is to 'increase parents' representation and bring them closer to the classroom in order to develop stronger home-school partnerships' (EQUIP-T MA, 2015, p. 3). The responsibilities, roles, and activities of PTPs are meant to be decided at school level based on each school's needs and priorities.

At endline, virtually all schools (99%) have a PTP, similar to the situation at midline (96%). However, this is a significant and large change from baseline, when only 14% of schools had a

parent–teacher group (defined as a group that includes parents and teachers and that meets regularly).

The provision of training for PTPs has increased significantly since midline but the training has failed to reach a large group of PTPs. At endline, 72% of PTPs have received training, provided by EQUIP-T, WEOs, head teachers, or other official Government agencies, on their responsibilities and roles, a significant and large increase from 39% at midline. The training shortfall will affect PTPs' capacity to support teachers, as well as general school activities and development, which is in turn likely to undermine any EQUIP-T impact on school and education quality coming through PTP activities.

Box 14: Has EQUIP-T provided training on SCs, PTPs, grants, and business plans for head teachers? (EQUIP-T input)

Head teachers are meant to receive EQUIP-T training to be able to support and engage with SCs, PTPs, and the wider community. During the last four years, out of the current head teachers, the majority have received such training, although the coverage varies by topic.

At endline, 76% of head teachers have received training on SC responsibilities and role; 60% on PTP responsibilities and role; 60% on the application and management of PTP grant 1; and 56% on how to develop a business plan and income generation. One reason that such large groups of head teachers have not received EQUIP-T training on these different topics is likely to be the high head teacher turnover (see in Box 11 Chapter 5). Regardless, this low training coverage is likely to affect head teachers' ability to support SCs and PTPs, and to engage with the community.

6.3.2 Has PTP capacity increased and are PTPs active? (EQUIP-T output)

The activeness of PTPs, based on the frequency of meetings and actions taken, is low.¹³⁰

On average, schools held fewer than two PTP meetings in 2017, and a sizeable minority of schools (18%) held no PTP meeting during the whole year.¹³¹ Only in 13% of schools did the PTP meet at least four times, which would roughly translate into quarterly meetings.¹³² Furthermore, among PTPs only 68% took some action during 2017 to improve education according to head teachers, which is similar (64%) to the reporting of Standards 1 to 3 teachers (also see section 6.3.3). Since close to all schools have a PTP, this suggests that many PTPs are not very active in terms of meetings or actions taken.

On average, PTPs are gender-balanced. PTPs are meant to consist of one father, one mother, and one teacher from each class, for a total of seven fathers, seven mothers, and seven teachers. At endline there are on average seven fathers and seven mothers on the PTP, and three female teachers and four male teachers. Among parents of Standard 3 pupils in the programme schools, 7% are members of a PTP.

There is still confusion among head teachers about the role of PTPs, compared to the role of SCs. At midline, the qualitative research found that head teachers were unclear about the difference between SCs and PTPs, and that they saw their roles as overlapping substantially (OPM, 2016b). This remains an issue at endline, with 18% of head teachers seeing the roles of

¹³⁰ In EQUIP-T's Annual Monitoring Survey 2017, p. viii, a 'PTP was considered active if it had been established, and if there was a recent activity reported by either the Head Teacher or members of the PTP'.

¹³¹ As PTPs are expected to carry out activities at a class level, there are no EQUIP-T prescribed requirements for the frequency of overall PTP meetings. However, PTPs have been eligible for two grants since 2016, which presumably require meetings to decide on spending plans, implementation plans and monitoring. Also the composition of the PTP must change annually, as the Standard 7 classes leave, and new Standard 1 classes start.

¹³² If meetings are relatively evenly spaced out over the school year.

PTP and SC as similar and 65% seeing them as only somewhat different, while 16% think the two roles are very different.

6.3.3 Have PTPs taken action to improve education, and are parents' engaged? (EQUIP-T intermediate outcome)

The proportion of schools where the PTP took some action to improve education in the last school year has increased significantly since midline. At midline, the PTP took action to improve education in 47% of schools, compared to 68% at endline, based on head teacher reporting.¹³³ At the same time, only 10% of parents of Standard 3 pupils report that the PTP took some action to improve education. This is likely related to the finding that at endline only 23% of parents of Standard 3 pupils are aware that a PTP exists at their child's school. Already at midline the evaluation's qualitative research found that many parents and teachers, unless they were part of the PTP, were not aware of its existence (OPM, 2016b). This lack of awareness, combined with the relatively low level of PTP activity, raises questions about PTPs' ability to help bridge the gap between schools and parents. Among parents who are not members of a PTP, only 9% have attended an information meeting about the PTP at their school, which to some extent may help explain their low awareness levels.

By far the most common action taken by PTPs to improve education in the last school year was to improve pupil attendance and punctuality. At endline, 46% of PTPs took such action, compared to 31% at midline (no significant difference). Despite a large proportion of PTPs taking action to improve pupil attendance, there was no significant change in overall attendance over the midline to endline period, and there was a significant but relatively small decline in attendance for boys (see Section 3.2 in Chapter 3). Among PTPs, 10% took action to improve school infrastructure; 9% took action to provide school feeding (which increased significantly from about 1% at midline); 8% took action to improve teacher attendance and punctuality; and 6% took action related to IGAs. 2% or less of PTPs took actions to provide extra-curricular activities, extra tuition classes, extra TLMs; for community members to assist in the classroom; fundraise; or improve pupil welfare. These findings indicate that raising pupil attendance is considered a top priority by PTPs. It is possible that the significant increase in PTPs providing school feeding is related to this, as it is associated with higher pupil attendance (Drake *et al.*, 2016). School feeding programmes are however, still very uncommon. Only 10% of head teachers reported that there is a programme in their school, and this coverage has not changed significantly since midline. It is possible that the increased involvement of PTPs in school feeding over the same period has potentially displaced other initiatives.¹³⁴

6.4 Findings: PTP and IGA grants

6.4.1 Has EQUIP-T provided PTP grants (EQUIP-T input) and have PTP grants been spent? (EQUIP-T output)

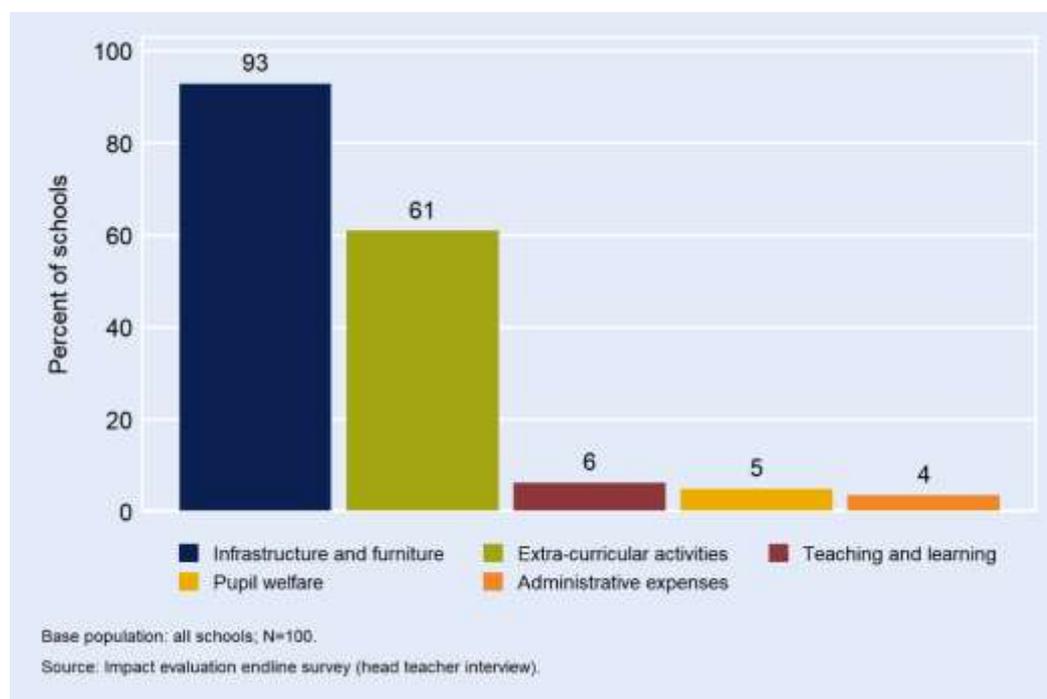
At endline, all schools have received PTP grant 1, and virtually all of them have received the correct amount (98%) and have also spent the grant (100%). The most common usage of PTP grant 1 is for school infrastructure and furniture (93% of schools) and extra-curricular activities

¹³³ Based on reporting by Standards 1 to 3 teachers who had been at the school at least one year, PTPs in 39% of schools took such action at midline, which rose significantly to 64% at endline.

¹³⁴ Head teachers were asked if there is a school feeding programme that has provided food for children in the last five days.

(61%), in line with the allowable uses of these grants. Much smaller proportions of schools spent the grant on teaching and learning (6%); pupil welfare (5%); and administrative expenses (4%).

Table 16: PTP grant 1 spending



6.4.2 Has EQUIP-T provided IGA training and grants (EQUIP-T input) and have IGAs started? (EQUIP-T output)

After midline EQUIP-T introduced training on how to develop business plans for head teachers / teachers and community leaders, to encourage communities and schools to work together to generate additional funds for school improvement through IGAs (EQUIP-T MA, 2017).

Almost all school–community groups (96%) have developed and submitted a business plan for IGAs to EQUIP-T at endline.¹³⁵ As the business plan includes specific materials it might be assumed that school–community groups would not have been able to develop and submit it without having received training on how to develop it. Under this assumption around 96% of school–community groups have received training on the business plans. Among all submitted business proposals (successful and unsuccessful), the most commonly proposed IGAs relate to livestock or livestock products (53%), and after that, agriculture or horticulture (22%), trading and sales (16%), and manufacturing and processing (6%).

The EQUIP-T target of providing IGA grants to 50% of all schools in the programme districts has essentially been achieved at endline. The IGA business plans were accepted for 49% of the programme schools, and close to all schools (98%) whose business plans were accepted received the IGA grant in its full amount.¹³⁶

Among the schools that received IGA grants, 88% have started at least some IGAs by endline, according to head teacher reporting. If the IGAs are successful in generating additional

¹³⁵ EQUIP-T worked with government officials, including regional government trade officials, to assess the submitted IGA business plans.

¹³⁶ In all cases the correct grant amount of TZS 1,500,000 was received.

income for schools, and this is used to address relevant school needs and priorities, this may not only contribute to improved school performance but also to sustainability after the programme ends. This leaves the question of how the 50% of schools that submitted IGA proposals that were unsuccessful will be able to meet funding shortfalls.

6.5 Findings: school notice boards

6.5.1 Has EQUIP-T provided school notice boards? (EQUIP-T input)

EQUIP-T has provided notice boards to schools to improve communication between schools and communities, and to increase the transparency of school finances and performance.¹³⁷ Schools are expected to place the notice boards in a publicly accessible space, and to display relevant information for school and community members.

Not all schools have yet received a notice board from EQUIP-T by endline. Based on head teacher reporting, 88% of schools received a notice board from EQUIP-T in the calendar years 2016 and 2017, which indicates a sizeable shortfall.¹³⁸

6.5.2 Are school notice boards publicly accessible and used? (EQUIP-T output)

The proportion of schools with notice boards that are publicly displayed on the school premises rose significantly from 49% at baseline to 72% at midline but have then declined significantly to 49% by endline (Figure 20).¹³⁹ The reason for the initial increase in schools with notice boards that are publicly displayed is likely to have been the provision of notice boards by EQUIP-T close to the midline survey. One possible reason for the subsequent decline might be that the provided notice boards were not particularly suitable for being displayed outdoors as their surface was made of fabric, without protective covers or a shelter.¹⁴⁰ Alternatively head teachers may have found it inconvenient to move them in and out of school each day or be concerned about them being stolen.

For schools with publicly displayed notice boards the most frequent type of information posted at endline is: academic results and teaching and learning information (56%) (a significant and large increase since baseline); SDP, budget, grants, and other financial information (35%) (also significantly up from baseline, but still a relatively low share given EQUIP-T's aim to improve financial transparency); community and school events (17%); and pupil and teacher attendance (10%). Less than 2% of the notice boards display information about JUU clubs or pupil welfare (see Chapter 7).¹⁴¹ This clearly shows that despite some improvement in the types of information displayed for schools that do have their notice boards in a public space on school premises, they do not use them to display a range of core school information.

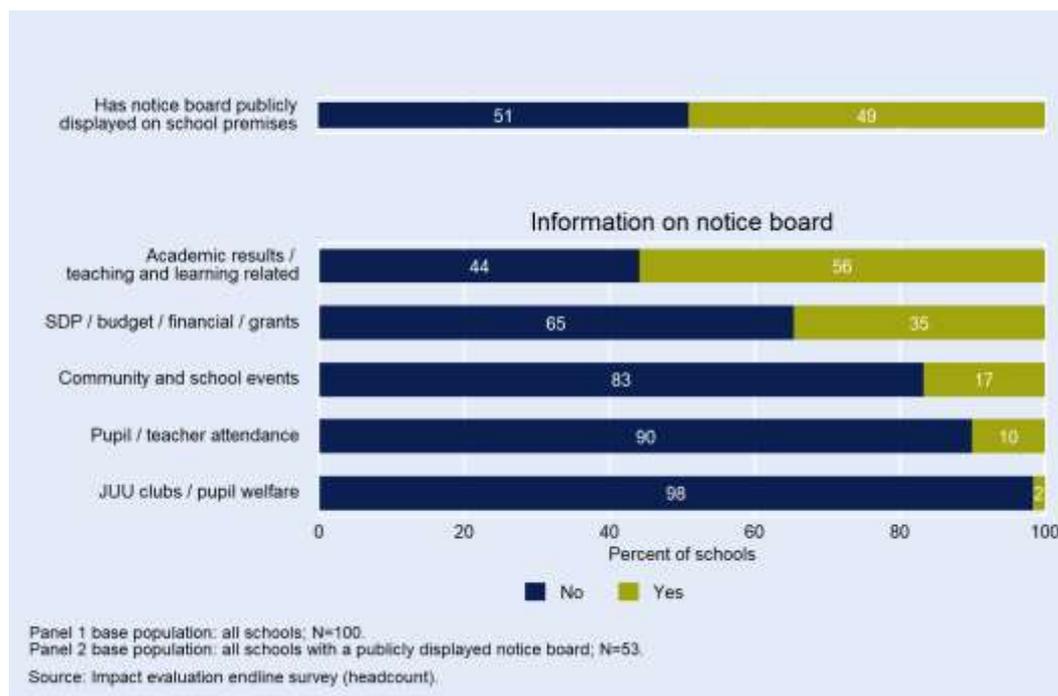
¹³⁷ At the time of the endline survey, implementation of the community score cards with school data from the new SIS meant to enable communities to hold schools to account and identify areas of improvement had not yet begun.

¹³⁸ This shortfall agrees with the figure in EQUIP-T's Annual Report 2015 (p. 3) which stated that the programme had delivered boards to 88% of target schools. However, there is a difference in reported timing between the EQUIP-T report, and head teachers' reporting. According to the EQUIP-T report, the boards were delivered towards the end of 2015 while head teachers could have misremembered this as 2016.

¹³⁹ At baseline and midline the survey enumerators were trained to check if schools had publicly displayed notice boards, whereas at endline they were similarly trained and the question (for enumerators to answer) explicitly asked about publicly displayed boards.

¹⁴⁰ The EQUIP-T MA in their written comments on the draft of this report (received 15 November 2018), pointed out that the noticeboards adhered to Government of Tanzania standards.

¹⁴¹ This category only exists at endline.

Figure 20: Accessibility and use of school notice boards at endline

6.5.3 Are parents aware of and reading school notice boards? (EQUIP-T intermediate outcome)

Only 29% of parents are aware that the school has a notice board. This is reflected in a mere 18% of parents having read the school notice board at least once during the first quarter of 2018. On average, parents who know the school has a notice board, checked and read it only 1.5 times during the school year. According to the evaluation's qualitative research at midline, most parents do not go the school regularly, and also many parents cannot read (OPM, 2016b). When comparing the awareness of noticeboards at endline between parents who are literate and those who are not, the difference is large and strongly significant (awareness is 33% among literate parents and 14% among parents who are illiterate). Another potential barrier to communication by noticeboard is the time taken for parents to reach the school from home, but the endline findings don't support this. Parents who live 45 minutes or more away from school are as equally likely as their counterparts who live within 45 minutes of school to know that the school has a noticeboard or to have checked it.

Overall the low awareness and reading of noticeboards strongly suggest that using notice boards to share school information with parents is unlikely to be effective as things stand.

6.6 Findings: CENAs

The aim of the CENAs under EQUIP-T is to encourage community engagement in schools while ensuring that schools take community needs and priorities into account. The CENAs are led by community facilitators who were trained by local CSOs before midline. The community facilitators are then meant to conduct the CENAs with community members and to develop action plans that feed into SDPs (EQUIP-T MA, 2017).

6.6.1 Was a CENA undertaken and actions taken based on it? (EQUIP-T input and intermediate outcome)

CENAs were conducted in 58% of the school communities during the period 2014 to 2017.

Among the communities that conducted a CENA over this period, 77% took some action to improve education based on it. Thus it appears that when communities undertake a CENA, a large proportion of them take some action based on it, and that the more challenging part is for communities to undertake a CENA to begin with. At endline, for the school communities that conducted a CENA during the last two years, the actions most frequently taken were: actions to improve school infrastructure (84%); actions to raise pupil attendance and punctuality (17%); school feeding (11%); monitoring teacher attendance (10%); and fundraising (8%).¹⁴²

EQUIP-T planned to complete all its CENA activities by February 2018, a month prior to the endline survey (EQUIP-T MA, 2017). Later this activity was extended and was still ongoing as of November 2018.¹⁴³

6.7 Has schools' interaction with all parents and the wider community improved? (EQUIP-T intermediate outcome)

Head teachers' ratings of community support for education is low, with only 23% considering it 'good' or 'very good' at endline, and this has not changed significantly over time. The relatively low share of communities that have conducted CENA activities by endline (see Section 6.6.1) may help explain the lack of change in this indicator together with parents typically not being aware of PTPs and not checking and reading school notice boards (see sections 6.3.3 and 6.5.3).

Schools do not communicate directly with a sizable share of parents about their child's academic progress during the school year. At endline, a large proportion of parents (40%) did not receive any written information or meet with a teacher about their child's academic progress during the previous school year. Meanwhile 9% only met with a teacher; 21% only received written information; and 29% both met with a teacher and received written information. This is despite parents considering academic progress important, judging by the main topics discussed in SC meetings and teacher–all parents meetings.

Schools are required by the Government to hold annual meetings of all parents and teachers as a basic communication and engagement mechanism (OPM, 2016b).¹⁴⁴ At endline and midline, close to all schools (96–97%) had held at least one such meeting, which is a significant increase from baseline (87%). The main topic discussed during the most recent meeting is academic progress (41%), a significant increase from baseline (28%) but not significantly different from midline (32%). The second most commonly discussed topic is pupil absenteeism, discipline, and dropout (20%). This is followed by school finance, including parental contributions (7% at endline and 12% at midline), which is significantly up from baseline, when it was not the main topic in any meeting. Meanwhile, infrastructure development is discussed in significantly fewer meetings at endline (4%) than at midline (9%), but is similar to baseline (5%). Teacher supervision and support is only discussed in 4% of meetings at endline, and teacher discipline is discussed in no meeting.

These findings suggest that SC meetings and annual meetings of teachers and all parents prioritise different topics. The former focus more on school finance (see section 6.2.2), whereas

¹⁴² The sample size for these estimates is small because only 18 schools conducted a CENA and took action based on it between midline and endline.

¹⁴³ This information came from the EQUIP-T MA as part of their feedback on the draft of this report.

¹⁴⁴ This refers to a meeting where all parents are invited by the head teachers; it does not include PTP meetings.

the latter focus more on academic progress and pupil absenteeism, discipline, and dropout. Neither of the two types of meetings prioritise teacher supervision, support, or discipline.

6.8 Summary of community 4A evidence

This section summarises the evaluation findings related to selected EQUIP-T community inputs, outputs, and intermediate outcomes, and Table 17 shows to what extent EQUIP-T has achieved these. Factors that may affect programme sustainability are also discussed as appropriate.

Table 17: Summary of evidence on community participation and accountability

Evaluation questions and assumptions related to the theory of change	Result at EL	Significant change	
		BL to EL	ML to EL
Has community participation in education and school-community communication improved in treatment schools, with changes consistent with the theory of change?			
Inputs Has EQUIP-T provided training for SCs and PTPs, grants, and school notice boards, and was a CENA undertaken?			
SC received training by EQUIP-T, WEO, head teacher, or other official govt. agency last two years (% schools)	87%	-	positive
PTP received training by EQUIP-T, WEO, head teacher, or other official govt. agency last two years (% schools)	72%	..	positive
Received PTP grant 1 (% schools)	100%
School and community developed and submitted business plan for IGA to EQUIP-T (% schools)	96%
Schools that received IGA grant as a proportion of the target number of schools (%) ¹⁴⁵	98%
Received school notice board from EQUIP-T by endline (% schools)	88%
Outputs Have SCs and PTPs become more active and are parents/communities more involved in school improvement?			
SC met last quarter (% schools)	69%	-	none
PTP exists (% schools)	99%	..	none
PTP met at least four times in 2017 (% schools)	13%
PTP grant 1 has been spent (% schools)	100%
Started at least some IGAs (% schools that received IGA grants)	88%
Notice board publicly displayed on school premises (% schools)	49%	..	negative
Intermediate outcomes Has engagement and communication between schools and all parents / wider communities increased?			
Did not meet with a teacher or receive written information about child's academic progress in 2017 (% parents of Std 3 pupils)	40%	-	-
Aware that PTP exists at school (% parents of Std 3 pupils)	23%	..	-
PTP took action to improve education (% schools)	68%	..	positive
Read notice board at least once during the first quarter of 2018 (% Std 3 parents)	18%	-	-
At least one meeting per year with head teacher, teachers, and all parents (% schools)	97%	positive	none
Notes: (1) .. means not applicable either because it is an EQUIP-T input, or because of the nature of the indicator. (2) - means trend data are not available.			

Fairly large groups of SCs and PTPs have still not received training from EQUIP-T, despite that all SCs and PTPs are meant to have received it at this point. This lack of training for some SCs and PTPs will weaken the path from training to increased SC and PTP capacity, and may reduce EQUIP-T's impact through the community component.

¹⁴⁵ The EQUIP-T target is for 50% of all schools in each programme LGA to receive IGA grants.

SCs are meant to meet at least quarterly, but still at endline a relatively large group of SCs met less frequently. This is similar to the situation at midline, and suggests that even if SCs receive training, it may not be enough to ensure their active engagement in school matters. For example, at midline several SCs indicated that they found it difficult to meet regularly because of a lack of incentives to attend meetings.

Among PTPs, a majority took some action to improve education during 2017 according to head teachers, and this is a significant increase since midline. Nevertheless, a large group of PTPs did not take any action at all during the year, reflective of the relatively low levels of PTP activeness. That not all PTPs have received training yet may be undermining the link between training, PTP capacity, and their ability to engage in school improvement, and may therefore weaken EQUIP-T's sustainability strategy. High turnover of head teachers and WEOs between midline and endline may also be a factor affecting PTP activity.

Despite the existence of PTPs at virtually all programme schools their activeness based on the frequency of meetings and actions taken is fairly low at endline, and they are not well known in the community. Only a small group of parents are aware that a PTP exists at their child's school, signalling that parents – other than those on the PTP – are not engaging with schools through the PTPs. The relatively low levels of PTP activeness and the lack of awareness of their existence are likely to undermine EQUIP-T's efforts to strengthen communities' and parents' engagement in education, something which is considered an important part of the sustainability strategy.

All schools have received PTP grants, and close to all school communities have submitted IGA business plans, and of those with successful proposals, nearly all have received IGA grants by endline.

Close to all programme schools have spent PTP grant 1 by endline. The grants were mainly spent on school infrastructure and furniture and extra-curricular activities, in line with the grants' allowed uses.

Among schools that have received IGA grants, the large majority have started at least some IGAs by endline. This is important from a sustainability aspect because if schools and communities successfully engage in IGAs, and if they are successful, this will help generate additional funds that could be used to continue effective school improvement activities started under EQUIP-T.

There is a shortfall in the distribution of notice boards to schools. All schools were expected to receive notice boards from EQUIP-T to increase school transparency and communication with parents and the community, but they did not, and this may undermine any effect of EQUIP-T that is meant to come through improved interaction between schools, parents, and communities.

Notice boards are publicly displayed on school premises in less than half of schools and are in most cases not used to display a range of core school information at endline. Between baseline and midline there was a significant increase in schools with publicly displayed notice boards, likely related to the provision of boards by EQUIP-T, but this has declined again by endline.

Most parents are not aware that the school has a notice board. This is reflected in only a small group of all parents having checked and read the school notice board at least once during the first quarter of the current school year. According to the evaluation's qualitative research at midline, most parents do not go the school, and also many parents cannot read. Taken together, these

findings strongly imply that using notice boards to increase transparency and share school information with parents is unlikely to be effective as things stand.

At both endline and midline close to all schools held annual meetings of teachers and all parents, a significant increase since baseline. At the same time, a large group of schools do not communicate with parents about individual pupils' academic progress during the school year. This is despite parents considering academic progress important, judging by the main topics discussed in SC meetings and annual teacher–all parents meetings.

7 Conducive learning environments for marginalised children, particularly for girls and children with disabilities

Following the extension of EQUIP-T in 2017, the programme has had a greater focus on promoting inclusive education, which has led to the introduction of the new component '4B'. Under this component, *Conducive learning environments for marginalised children, particularly for girls and children with disabilities*, EQUIP-T is testing and scaling up approaches to support the education of girls, children with disabilities, and other marginalised groups. These measures are anticipated to promote inclusive education, thereby improving pupil learning achievement, narrowing learning gaps, and increasing retention and transition rates to secondary schools, especially for marginalised groups and girls in particular.

This new focus over the final years of the programme strongly supports the Government of Tanzania's new National Strategy for Inclusive Education (2018–2021), which aims to provide quality and equitable learning opportunities for all children, adolescents, and youth, with a focus on excluded and marginalised groups (MOEST, 2017). This is to be achieved through promoting an 'inclusive education system that removes the barriers limiting the participation and achievement of all learners, that respects diverse needs, abilities and characteristics and that eliminates all forms of discrimination in the learning environment.' (MOEST, 2017 p. 3).

Box 15 describes some of the key concepts for this chapter.

Box 15: Key concepts for EQUIP-T conducive learning environments component

Inclusive education ensures that every child receives a high quality education, regardless of attributes such as gender, physical and social attributes, intellectual status, linguistic background, or special needs (UN Children's Fund (UNICEF), 2009). Inclusion can be viewed as the process of addressing and responding to the diverse needs of all learners, and may involve changes and modifications in content, approaches, structures, and strategies, to fulfil the responsibility of the regular education system to educate all children (UNESCO, 2005).

Marginalised or vulnerable children are children who, due to their gender, physical characteristics, or socioeconomic background, are at a disadvantage in education compared with other children. This includes groups such as girls, children with disabilities, orphans, children who live in poverty, children who do not speak Kiswahili as their first language, children who live in remote communities, and working children.

Pupil welfare includes any policies, actions, or clubs that promote the welfare of pupils in the school and that are not related to academic or extra-curricular activities. This covers the health, safety, and well-being of pupils. Examples include menstruation support for girls, addressing female genital mutilation, rights to education campaigns, counselling, encouraging positive discipline rather than corporal punishment, etc.

JUU clubs are non-academic and non-extracurricular school clubs for pupils in upper standards that 'provide a platform for children, especially girls, to receive guidance on managing the barriers to education, to discuss their opinions and to act as change-makers in their school and communities.' (EQUIP-T MA, 2017 p. 30) JUU stands for '*Jiamini Uwezo Unao*' in Kiswahili, which is a motivational slogan that translates as 'Be confident, you can'.

This chapter begins by providing an overview of this component's theory of change, programme implementation since midline, and changes expected by the programme as a result of its activities. This is followed by contextual evidence on the profile of vulnerability of pupils in the programme schools. The quantitative findings on conducive learning environments are then presented, structured according to the theory of change. Finally, the chapter summarises key findings, and

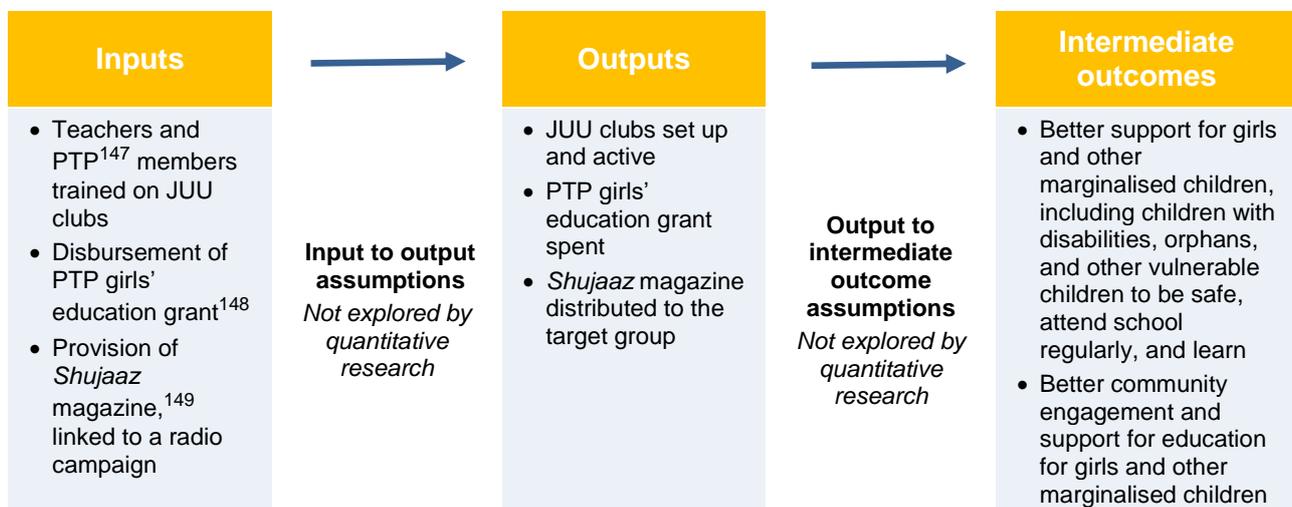
maps out the extent to which different inputs have been successful in achieving their intended outputs and intermediate outcomes.

7.1 EQUIP-T component 4B: theory of change, implementation, and expectations of change

7.1.1 Theory of change

Figure 21 shows the inputs, outputs, and intermediate outcomes in the main results chain for EQUIP-T component 4B, which are covered by the quantitative research. It excludes theory of change assumptions, as these are not covered by the quantitative research for this particular component.

Figure 21: Main results chain for EQUIP-T component 4B conducive learning environments¹⁴⁶



In addition to the inputs depicted in Figure 21, there are other existing initiatives aimed at promoting gender inclusion, and, to a lesser extent, social inclusion, in all components of the programme – including, for example, teacher in-service training on gender-responsive pedagogy, as well as new initiatives that were yet to be rolled out at the time of the endline survey.

The quantitative research provides partial evidence against this component. This is because most of the interventions will have had insufficient implementation time to be fully evaluated in 2018; and these types of interventions related to empowerment, and tackling cultural norms and taboo issues, are more suited to qualitative research methods.

¹⁴⁶ See Annex C for the full endline evaluation matrix, which includes the theory of change assumptions for component 4B that are not covered by the quantitative research.

¹⁴⁷ See Box 13 in Chapter 6 for a definition of PTPs.

¹⁴⁸ The grant is meant to be used to improve the attendance, retention, learning, and welfare of girls. However, the EQUIP-T MA clarified that children with disabilities and other marginalised children could also benefit from the grant.

¹⁴⁹ *Shujaaz* magazine is a colourful comic book-style publication that focuses on a young girl and the challenges she faces to attain education.

7.1.2 Implementation since midline¹⁵⁰

Under component 4B, three main activities have been delivered since midline: training to enable schools to set up JUU clubs for pupils in upper standards; provision of a second PTP grant for activities to support girls' education; and distribution of the *Shujaaz* magazine to schools and communities as part of a behaviour change campaign related to girls' education (OPM, 2018).

The programme is planning a number of other activities under this component that are under development or had not yet reached the schools by the time of the endline survey (see Box 16).

Box 16: Component 4B activities planned from mid-2018 onwards

These include: a **gender-responsive toolkit** that will address issues of gender-based violence and menstrual hygiene management for girls; an **activity pack for JUU clubs**; a **pilot initiative to address child protection** issues, which may include suggestion boxes for pupils to anonymously report issues (the endline survey finds that suggestion boxes already exist in some schools); **female genital mutilation initiatives** in three regions; **additional training to teachers** on general pedagogy, including support **on inclusive education and supporting pupils with disabilities**; and a **secondary school readiness programme for girls** to improve girls' transition from primary to secondary school (EQUIP-T MA, 2017).

For more details on EQUIP-T implementation see Annex B.3.

7.1.3 Expected changes

EQUIP-T staff expect three main changes to result from the activities implemented since midline under this component: active JUU clubs; increased knowledge and openness of teachers and pupils on barriers to girls' education, and barriers to inclusive education, and examples of how to tackle these; and increased retention rates for girls in school (OPM, 2018).

The programme is also monitoring a number of indicators under this component that are stated in its logframe. These relate to PTP support to improve inclusion or child protection; establishment and activeness of school pupil welfare clubs; pupil welfare club support to girls' welfare or gender equity; and existence of suggestion boxes in schools (EQUIP-T MA, 2017).¹⁵¹ For more details on the logframe indicators see Annex B.6.

7.2 Profile of vulnerability in programme schools

This section provides a profile of Standard 3 pupils and their different dimensions of vulnerability.

Sizeable shares of Standard 3 pupils in the programme schools suffer from multiple disadvantages that are considered among the key barriers to education. Poverty, gender, ethnic and linguistic background, disability, geographical location, and livelihood can all act as barriers to education (UNESCO, 2015). About 40% of pupils come from households living below the poverty line, and more than 70% of pupils did not eat any food before coming to school on the day of the survey. These shares have slightly increased over time, although the change is weakly significant. Particularly striking is that Kiswahili, the main language of instruction, is not spoken as a main language at home for over three-quarters of pupils. A sizeable minority of pupils (31%)

¹⁵⁰ As this is a new component that has been introduced by EQUIP-T since midline, there were no interventions by the programme between baseline and midline.

¹⁵¹ Note that while this evaluation measures activeness of JUU clubs and support provided by JUU clubs and PTPs in relation to the inclusion and welfare of pupils, these indicators are not defined in the same way as the logframe indicators.

come from remote communities and take 45 minutes or longer to get to school every morning. Physical disabilities are quite prevalent: 16% of pupils have some or a lot of difficulty seeing, hearing, walking, remembering, or concentrating (as self-reported by the pupils). There is also a notable share of pupils who do paid or unpaid work outside their household (18%), and this share has increased significantly since midline. Being in school over-age, another dimension of vulnerability, is prevalent among 46% of the pupils. Just over half of Standard 3 pupils (51%) are female, which means that enrolment in schooling is gender-balanced, but enrolled girls often face substantial additional obstacles to learning, both in school and at home.

Regular school attendance is worse for certain groups of vulnerable pupils (Table 18). The biggest gap is between pupils from poorer households and pupils from richer ones. On average, pupils from poorer households were absent for 16% of the school days during the period January to March 2018, while their peers from richer households were absent, on average, for 11% of the days, and this gap is strongly significant. Distance to school is another factor: pupils who live within 45 minutes of the school were absent, on average, for 12% of the school days during January to March, while those who live 45 minutes or more away from the school were absent for 15% of the days. Over-age pupils and those who do not speak Kiswahili at home are also significantly more likely to be absent from school than their peers.

Table 18: Pupil absenteeism from school by certain vulnerability dimensions

	Endline							
	Poorer	Richer	Travel far	Travel close	Local language	Kiswahili	Over-age	Not over-age
Proportion of days absent in Jan–Mar 2018 (mean % days)	16.0	11.2***	14.7	11.5**	13.6	10.6**	14.2	11.6**
<i>N</i>	563	880	384	873	1,167	279	645	778
Sources: Evaluation endline survey (school records).								
Notes: (1) Asterisks indicate statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. (2) This is the mean percent of days absent over all tested Standard 3 pupils.								

Linguistic background and poverty status continue to be markers of disadvantage in educational attainment. Pupils who do not speak Kiswahili at home are significantly more likely to be in the bottom performance band in both Kiswahili and maths, and significantly less likely to be in the top performance band for maths (see Table 5 in Chapter 3). Similarly, pupils from poorer households have far worse literacy and numeracy skills on average than pupils from richer households. On the other hand, girls have made significant gains in learning outcomes compared to boys: on average, they now significantly outperform boys in Kiswahili and have considerably reduced the performance gap in maths since baseline (see Table 4 in Chapter 3).

It is important to note, however, that pupils' home language and poverty status are correlated with each other as well as with other indicators of education disadvantage such as living far from school, not eating food in the morning and working outside the household. In order to isolate the correlations between these different characteristics and pupil learning outcomes, a simple regression analysis was carried out. This showed that home language is significantly and independently correlated with pupil learning – that is pupils who don't speak Kiswahili as their native language are more likely to have lower performance in both maths and Kiswahili skills. Poverty status, however, was only found to have a negative but weakly significant correlation with the proportion of pupils in the bottom performance band for Kiswahili. Other factors that were found to be independently correlated with learning outcomes are pupil age (positive effect); pre-school attendance (positive effect); having difficult remembering or concentrating (negative effect); having

support for education at home (positive effect); and being absent from school (negative effect). Additionally, being female was found to be positively correlated with Kiswahili performance once other background factors were controlled for.¹⁵²

While the vulnerability of many pupils in the programme schools is expected, given that the programme operates in some of the most remote and economically disadvantaged districts in the country, this underscores the necessity of promoting inclusive education in these schools. This is particularly the case given the evidence showing that some groups of vulnerable pupils have worse attendance rates and learning outcomes than their peers.

7.3 Findings

This section presents quantitative evidence from the evaluation baseline, midline, and endline surveys to assess whether key inputs under component 4B have been delivered as planned, and if outputs and intermediate outcomes have been achieved.

7.3.1 Have schools received the various inputs under this component as intended? (EQUIP-T input)

The various interventions under this component related to the JUU clubs, PTP girls' education grants, and *Shujaaz* magazines have been delivered to schools largely as intended. Over 90% of schools report receiving training in 2016 or 2017 on setting up and running a JUU club. The PTP girls' education grant was received by 89% of schools, all of which received it in full.¹⁵³ Additionally, 87% of schools received copies of the *Shujaaz* magazine in 2016 or 2017. In 85% of schools with a JUU club the club has access to copies of the *Shujaaz* magazine.

7.3.2 Are JUU clubs established, active, and supporting marginalised children? (EQUIP-T output and intermediate outcome)

JUU clubs have been established in the large majority of schools (91%). Of all the schools with a JUU club, all but one formed the club in 2017 or earlier. The composition of the JUU clubs are, on average, gender-balanced and JUU clubs are typically made up of about 16 male pupils and 17 female pupils.

The majority of JUU clubs have been active over the past year. On average, JUU clubs met five times in 2017 and 79% of schools with a JUU club report that the club carried out at least one activity in 2017.

The most common activities that were undertaken by JUU clubs in 2017 were environmental (Figure 22). Of all schools with an active JUU club, 52% report that the club carried out activities related to the school and learning environment in 2017. For example, this could be carrying out a campaign in the community to advocate for the construction of new classrooms or new latrines for girls; or working on building or improving the school garden.¹⁵⁴ JUU clubs have implemented activities addressing issues related to the health and hygiene of pupils, and rights to education in about 31% of schools. Only 13% of active clubs carried out activities related to pupils' attendance and punctuality; for example, these could be home visit campaigns aimed at advising children who

¹⁵² The results from the simple regression analysis are available on request.

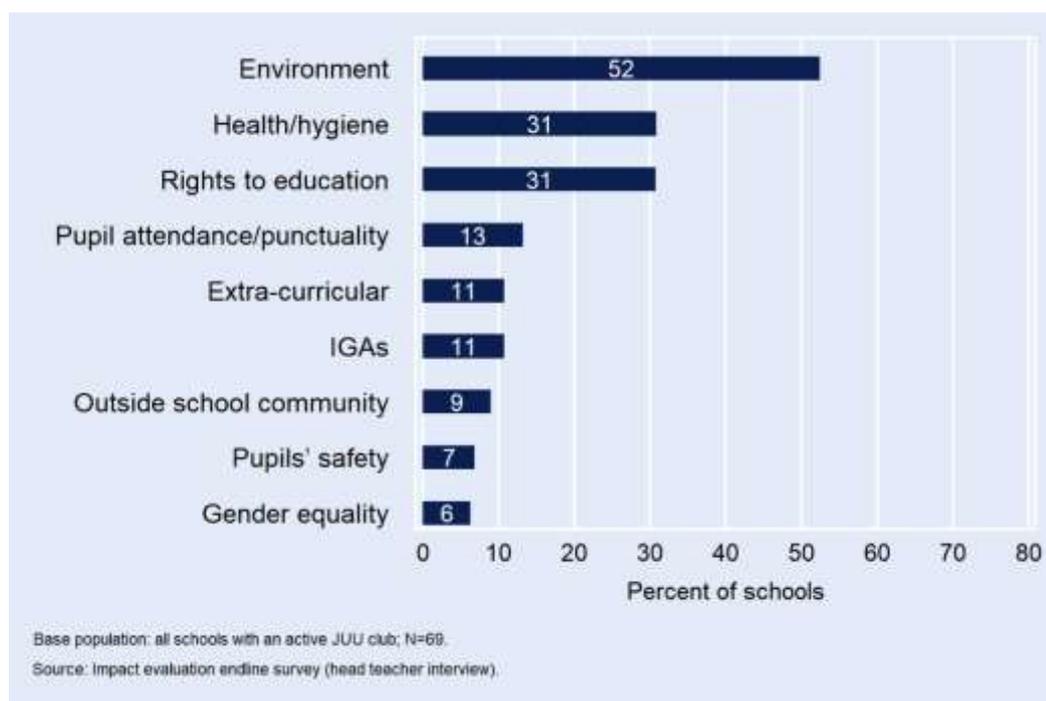
¹⁵³ The expected PTP girls' education grant is TZS 550,000 per school.

¹⁵⁴ These examples are generic, and do not come from the evaluation survey.

are not in school, and their parents, of the importance of sending children to school (EQUIP-T MA, 2017). Less than 10% carried out activities related to gender equality or pupils' safety.

A minority of JUU clubs have targeted groups of vulnerable pupils in their activities in the past year. Of all schools with an active JUU club, 84% report that none of the activities of the JUU club in 2017 were targeted at groups of vulnerable pupils. Only 7% report that some of the club activities were targeted at girls, and similarly 7% report that club activities were targeted at pupils with parents who are not interested in education.

Figure 22: Types of activities carried out by JUU clubs in 2017



7.3.3 Has the PTP girls' education grant been spent, and on what? (EQUIP-T output and intermediate outcome)

Only a quarter of schools have spent the PTP girls' education grant by the time of the endline survey. This is a very low share given when the grant was received: 81% of schools that received the grant in 2017 had still not spent it by April or May 2018. It is not clear why the majority of schools have not spent the grant, particularly as PTPs were meant to have received direction from EQUIP-T on how best to use the grant (EQUIP-T MA, 2017).¹⁵⁵

The most common expenditures made using the PTP girls' education grant were related to pupil welfare and infrastructure and furniture. The PTP girls' education grant was meant to be used to improve the attendance, retention, learning, and welfare of marginalised girls, as well as possibly children with disabilities and other marginalised children (EQUIP-T MA, 2017). Of all schools which have spent the grant, 66% spent all or some of it on pupil welfare activities, 66% spent it on infrastructure and furniture, and 35% spent it on extra-curricular activities. Additionally, 70% of schools that report spending the grant state that the expenditures made using the grant were targeted at girls, while the remaining 30% report that the grant expenditures were not targeted at any group of vulnerable pupils. Therefore, it is possible that some of the expenditures

¹⁵⁵ EQUIP-T produced a manual with guidance on the use of PTP grants to address girls' education issues that head teachers were supposed to use to train PTPs.

on infrastructure and furniture could have been to build toilets for girls, purchase desks for girls, or invest in other infrastructure that would promote the learning and welfare of girls at the school. It is important to note, however, that these shares correspond to the small sample size of schools that have spent the grant.

7.3.4 Have inputs related to child protection and positive behaviour management been received and are schools promoting child protection and anti-violence? (EQUIP-T input, output, and intermediate outcome)

Inputs, outputs, and outcomes related to positive behaviour management are classified as part of the theory of change for component 1A, *Teacher performance* (Figure 7). However, they are discussed in this chapter as they also relate to inclusive education.

As part of a national campaign to promote positive and safe learning environments, EQUIP-T provided schools with posters and activities related to teachers' professional code of ethics and conduct. The main goal was to encourage positive behaviour management in the classroom and reduce/eliminate the use of corporal punishment.

Just over a quarter (26%) of schools report receiving the posters on positive and safe learning environment in 2016 or 2017. Furthermore, in schools that received the posters, the posters were displayed on the classroom walls in only 5% of lessons observed. This low share could in part be explained by the fact that these posters were mainly meant to be displayed in teacher staff rooms as opposed to classrooms.

The use of corporal punishment in schools is almost universal. Of all interviewed parents of Standard 3 pupils, 96% report that their child's school practices corporal punishment, while 98% of that share report that their child was beaten at school as a punishment in 2017.¹⁵⁶ While the use of corporal punishment in schools is lawful in Tanzania,¹⁵⁷ it has been linked to worse learning outcomes in several contexts (Gershoff, 2017; Ogando Portela and Pells, 2015). Furthermore, the qualitative research at midline found that corporal punishment was the thing that pupils disliked the most about school (OPM, 2017a).

A suggestion box pupils can use to raise child protection concerns anonymously is available in 23% of schools. This does not necessarily point to low coverage, as the programme had not fully implemented this activity by the time of the endline survey. However, in the few schools where these boxes exist, they are located in a space that makes it difficult for pupils to access them. In 53% of schools that have a suggestion box, the box is displayed in an open space, which poses barriers to pupils dropping an issue into the box unnoticed, while in 22% of the schools the suggestion box does not have a permanent location, which results in pupils not always knowing where the box is placed.

7.3.5 Are teachers using inclusive teaching practices in the classroom to support marginalised children? (EQUIP-T intermediate outcome)

Gender-responsive pedagogy

Teachers' interactions with pupils in the classroom are significantly more gender-balanced at endline than they were at baseline. In 68% of lessons observed at endline, teachers engage

¹⁵⁶ The survey does not collect information on whether corporal punishment was used by teachers in the classroom or outside the classroom for disciplinary offences (for example, by the head teacher).

¹⁵⁷ Source: November 2013, CRC/C/TZA/3-5, Third-fifth state party report, para. 88.

proportionally with boys and girls present in the classroom, compared with 54% of lessons at baseline (Table 7 in Chapter 4). However, there are still about 19% of lessons where teachers engage more with boys, compared to 13% of lessons where teachers engage more with girls. Of all Standards 1 and 2 teachers, 72% report having completed the gender-responsive pedagogy module of the EQUIP-T in-service training; however, there was no significant change in the gender balance of teacher interactions between midline and endline.

Identification of and support to pupils with special learning needs

Almost all teachers (96%) report that as part of their teaching of Standards 1 to 3 they notice groups of pupils who find learning particularly difficult. Pupils from poor households are identified as having learning difficulties by 46% of teachers, followed by pupils who do not speak Kiswahili at home (45%), pupils who are regularly absent (40%), pupils whose parents are not interested in education (32%), pupils who live far from school (26%), pupils who have not attended pre-school (14%), and pupils with disabilities (13%).

Virtually all teachers (99%) report that they are able to help those groups of pupils that they identify as having learning difficulties. The most common strategies to support these pupils reported by teachers are: giving extra tuition classes (44% of teachers), which are classes that teachers provide to pupils outside of teaching hours either for free or for a fee; talking to pupils' parents (43%); grouping pupils together during the lesson (30%); giving more exercises and work (25%); and repeating topics until pupils understand (20%). Only 15% of teachers report a more general strategy of ensuring the active engagement of these groups of pupils in the lessons, while a mere 5% report that they adapt materials and teaching to the appropriate level.

There have been zero to moderate changes since midline in teachers' reporting of their identification of and support to pupils with extra learning needs. The changes are that more teachers at endline report pupils from poor households and pupils with parents not interested in education as having learning difficulties, while fewer teachers report boys finding learning difficult.¹⁵⁸ With respect to strategies to support these pupils, more teachers at endline report that they talk to the pupils' parents, repeat topics, and give more work and exercises, while fewer teachers report giving extra tuition classes. The moderate change since midline is unsurprising given that the EQUIP-T in-service training delivered to teachers by endline has not focused on inclusive education, beyond gender-responsive pedagogy. The main interventions under this area are yet to be rolled out.

Language of instruction

The language of instruction is of considerable importance in multilingual countries such as Tanzania, where many pupils do not speak the official language of instruction as a main language at home, which means that they have far fewer opportunities to practise speaking the language.

All Standards 1 to 3 teachers at endline report that Kiswahili is the main language of instruction used in their classrooms, which is expected given that it is the official language of instruction. Additionally, 92% of teachers always use Kiswahili to speak with pupils outside the classroom, and this has not changed significantly since midline.

¹⁵⁸ While the results in Volume II, Chapter 5 show that the share of teachers reporting pupils who are regularly absent and pupils who live far from school as having learning difficulties has increased significantly since midline, it is worth noting that these results are not directly comparable due to a change in the administration of this question between midline and endline – that is, these two categories were added to the list of categories in the questionnaire at endline, while at midline they were reported by teachers under the *other* category.

However, the vast majority of pupils (81%) come from households that do not speak Kiswahili as their first language, the same as at midline. This large misalignment between home and school language is likely to be a limiting factor in the learning process for those pupils who do not speak the main language of instruction at home (UNESCO, 2015). This is supported by this evaluation, which finds that this group of pupils have far worse literacy and numeracy skills on average than their peers who speak Kiswahili at home, and almost half of teachers (45%) find that this group of pupils faces learning difficulties.

Despite pupils who do not speak Kiswahili at home falling behind their peers in learning outcomes and almost half of teachers recognising that these pupils have learning difficulties, very few teachers are able to mitigate the language barriers in their classrooms. Teachers were observed to switch between Kiswahili and a vernacular language during the lesson in only 2% of observed lessons (unchanged since baseline and midline).¹⁵⁹ This is not surprising, given that this policy is discouraged in classrooms in Tanzania. However, there is no evidence that teachers are adopting other techniques to mitigate the language barriers as teachers were observed to provide extra support to non-native Kiswahili speaking pupils in only 3% of lessons observed (similar to the situation at midline).

There is likely to be multiple reasons why teachers are not providing additional support to pupils who don't speak Kiswahili as a home language, one of which is the shortage of teachers who speak the local languages of the pupils – only 7% of teachers speak a language other than Kiswahili at home and only 31% of pupils report that their teachers speak their local language (the latter share, however, has increased by 10 percentage points since midline). Another factor may be a lack of Government policy and guidelines on mitigating language barriers in primary classrooms.¹⁶⁰ EQUIP-T's main strategy to support children who do not speak Kiswahili at home is by improving their school readiness through the SRP (see Box 5 in Chapter 3), and via supporting national approaches to pre-primary education. However, there are clearly still major constraints on these pupils in primary schools.¹⁶¹

7.3.6 Are there inclusive strategies at the school level that support marginalised children? (EQUIP-T intermediate outcome)

Initiatives to improve pupil welfare and support marginalised groups of pupils at the school level have been very limited. Only 13% of schools have an SDP that has specific strategies to improve learning and welfare for girls and other vulnerable pupils. Virtually none of the head teachers at endline report taking action in 2017 to improve pupil welfare, while only 12% report that some of their actions in 2017 were to improve education for groups of vulnerable pupils.¹⁶² The WEOs' visits to schools have also not included any discussions or actions surrounding the welfare of pupils or inclusive education for girls and other vulnerable pupils.¹⁶³

¹⁵⁹ On the other hand, 35% of Standards 1 to 3 teachers self-report that they switch between Kiswahili and a vernacular language during the lesson.

¹⁶⁰ The new national inclusive education strategy only mentions language barriers in relation to actions to encourage pre-primary education and SRP.

¹⁶¹ At endline less than 1% of parents of Standard 3 pupils report that their child had attended EQUIP-T SRP. This is not surprising given that the SRP started in October 2015 only as a pilot phase and in a quarter of the EQUIP-T districts.

¹⁶² This share is higher when reported by interviewed Standards 1 to 3 teachers: 35% of teachers report that the head teacher took some action in 2017 to improve education for groups of vulnerable pupils, mostly for pupils who live far from school, pupils from poor households, and pupils who are regularly absent.

¹⁶³ None of these indicators were measured at baseline or midline, hence it is not possible to look at their trend over time.

Such initiatives have also been quite limited at the community level. While 46% of head teachers report that the PTP carried out activities in 2017 to improve pupil attendance and punctuality, and 9% report that the PTP took action on school feeding, none report that the PTP took any other action in 2017 to improve the welfare of pupils at the school. When it comes to supporting marginalised groups of children, only 29% of head teachers report that some of the PTP's actions in 2017 were targeted at vulnerable pupils, most commonly pupils who live far from school and pupils who are regularly absent. At the last all parents–teachers meeting held at the school, pupil welfare was the main topic of discussion in only 14% of the schools, and this was mostly on school feeding.

Very few schools have a school feeding programme, which has negative implications for pupils' concentration and participation in the learning process. The majority of pupils (71%) did not eat food before coming to school on the day of the survey, and this share has increased since midline. Compounding this in some cases, about a third of pupils take 45 minutes or longer to reach the school every morning, meaning that these pupils may well arrive at school already tired and many of them have not had food in the morning. However, only 10% of schools have a feeding programme that had provided food to pupils in the past five school days, and this share is not significantly different at midline. This creates a hunger problem which makes it more difficult for pupils to concentrate and learn properly (Pivik et al, 2012 and Hoyland et al, 2009). EQUIP-T does not have specific interventions to promote school feeding.

The overwhelming majority of schools (95%) have a gender coordinator—a position proposed by EQUIP-T. The intention is that this person is in charge of coordinating gender issues in the school, such as policies that promote the welfare and inclusion of girls. This could include listening to gender equity-related concerns from pupils, ensuring that seating in the classrooms is gender-balanced (that is, boys and girls sit in all areas of the classroom and share benches), and supporting girls with their menstrual hygiene management. The gender coordinator is not a formal position, but the teacher chosen to undertake this role is informally elected by all teachers in the school. This role may also include, but does not necessarily have to include, managing and counselling the JUU club. Most gender coordinators (83%) have attended specific gender training, such as on gender-responsive pedagogy or gender-inclusive environments, in 2016 or 2017.

A very small share of schools (4%) have special classes – that is, classes for pupils with special learning needs, such as pupils with disabilities.

7.4 Summary of findings on conducive learning environments

This section summarises the evaluation evidence on EQUIP-T component 4B, conducive learning environments. Table 19 provides an overview of the extent to which EQUIP-T has achieved its intended inputs, outputs, and intermediate outcomes.

Table 19: Summary of evidence on conducive learning environments for marginalised children

Evaluation questions and assumptions related to the theory of change	Result at EL	Significant change	
		BL to EL	ML to EL
Is there a conducive learning environment for marginalised children, particularly for girls and children with disabilities, and are changes as expected in the theory of change?			
Inputs Has EQUIP-T provided training on JUU clubs, <i>Shujaaz</i> magazine, and PTP girls' education grant?			
Received training on setting up and running a JUU club in 2016 or 2017 (% schools)	93%
Received the PTP girls' education grant (% schools)	89%
Received copies of the <i>Shujaaz</i> magazine in 2016 or 2017 (% schools)	87%

Outputs Are JUU clubs established and active and has the PTP girls' education grant been spent?			
JUU club has been established (% schools)	91%
JUU club carried out any activity in 2017 (% schools with a JUU club)	79%
Spent any of the PTP girls' education grant (% schools)	25%
School has a teacher responsible for coordinating gender issues at school (% schools)	95%	-	-
Intermediate outcomes Has there been better support for marginalised children to be safe, attend school regularly, and learn?			
JUU club activities in 2017 were related to pupil attendance, health/hygiene, safety, rights to education, or gender equality (% schools with an active JUU club)	64%
Some of the JUU club activities in 2017 were for particular groups of vulnerable pupils (% schools with active JUU club)	16%
Teachers' interactions with pupils in the classroom are gender-balanced (% lessons)	68%	positive	none
Teachers provide extra support to non-native Kiswahili speaking pupils during the lesson (% lessons)	3%	-	none
School does not practice corporal punishment (% parents of Std. 3 pupils)	4%	-	-
SDP includes specific strategies to improve learning or welfare for girls or other vulnerable groups of pupils (% schools)	13%	-	-
Head teacher took action in the past year to improve pupil welfare or improve education for groups of vulnerable pupils (% head teachers)	12%	-	-
PTP took action in the past year to improve learning or welfare for groups of vulnerable pupils (% schools)	29%	..	-
Main topic of discussion at the last all parents–teachers meeting was on pupil welfare (% schools)	14%	-	-
School has a feeding programme that provided food to pupils in past five days (% schools)	10%	-	none
Gender coordinator attended specific gender training in 2016 or 2017 (% schools with a coordinator)	83%	-	-
Notes: (1) .. means not applicable either because it is an EQUIP-T input, or because of the nature of the indicator. (2) - means trend data are not available.			

The vast majority of schools received the PTP girls' education grant, copies of the *Shujaaz* magazine, and training on the JUU clubs. On the other hand, posters on positive and safe learning environments, which are classified as an intervention under EQUIP-T component 1A but discussed in this chapter, have been received by only a quarter of schools.

JUU clubs are established in the overwhelming majority of schools and they are active in a large share of them. They have met a few times in the past school year and the majority have carried out an activity.

Despite the majority of schools receiving the PTP girls' education grant, only a quarter had spent any money from the grant by the time of the endline survey. It is unclear why this share is this low.

Although some JUU clubs have carried out pupil welfare activities that are related, for example, to gender equity, health and hygiene, child protection and rights to education issues, the primary focus in their early stages of operation was on the classroom and learning environment. Only a minority of schools report that the JUU club activities in the last school year were targeted at groups of vulnerable pupils.

Inclusive teaching practices in the classroom have been mixed. While teachers' interactions with pupils in the classroom are significantly more gender-balanced at endline than at baseline, there remains a sizeable minority of lessons where teachers interact more with boys than girls. Nearly all teachers report that they notice groups of pupils with learning difficulties in their teaching of Standards 1 to 3, and are able to support them. However, in virtually none of the lessons observed were teachers providing extra support to pupils who do not speak Kiswahili at home, despite the difficulties in learning that this group has been shown to face.

In general, initiatives to improve pupil welfare and support marginalised groups of pupils at the school and community level are quite limited. Very few schools include strategies in their SDP to improve learning or welfare for girls or other groups of vulnerable pupils, and, similarly, a minority report that the head teacher took action in 2017 to improve pupil welfare or improve education for certain groups of vulnerable pupils. PTPs have also largely not taken action to improve welfare or education for vulnerable pupils. Despite a stark and large share of pupils who do not eat any food before coming to school in the morning and who come from poor households, only a minority of schools have a feeding programme. Corporal punishment is practised by almost all of the schools, and while this is legal it is at odds with the promotion of positive and safe learning environments. The overwhelming majority of schools have a teacher that is responsible for coordinating gender issues, and most of these teachers have received specific gender training in the past two years.

PART C: Conclusions, recommendations, and lessons



8 Conclusions, recommendations, and lessons

This chapter brings together the findings presented in Part B above, to draw conclusions about the impact, effectiveness, relevance, and sustainability of EQUIP-T, based on evidence from this quantitative endline evaluation.¹⁶⁴ The report is intended to inform further adjustments to the programme before it finishes in January 2020, as well as to promote accountability and lesson learning for DFID and the Government of Tanzania. Thus, following the summary of findings, this chapter sets out recommendations that are grouped according to those useful for the EQUIP-T MA, and for the Government of Tanzania. The chapter identifies a number of lessons learned which would be valuable to readers outside of the Tanzania context involved in designing and implementing education programmes with similar objectives. Finally, the chapter introduces the overarching research questions (emerging from the quantitative analysis) that will be prioritised in the qualitative research in 2019.

8.1 Conclusions on impact

The overall goal of EQUIP-T is to achieve better learning outcomes at primary level, especially for girls. This impact evaluation finds that the programme has had a positive impact on both literacy and numeracy skills for pupils, compared with pupils in control schools. The impact on Kiswahili skills has resulted in more Standard 3 pupils in the top achievement band, and fewer in the bottom band, than would have happened in the absence of EQUIP-T. Between baseline and midline the programme brought pupils out of the bottom band, and between midline and endline the programme moved more children into the top band for Kiswahili. In maths, EQUIP-T had an impact between midline and endline in maintaining maths scores; in the absence of the programme, maths performance would have dropped. These results clearly show that EQUIP-T has had some success in achieving its overall impact-level goal.

The impact evaluation measures programme impact on two other indicators. The first, relating to teacher performance, is school and classroom absenteeism of teachers. There is weak evidence of a positive impact on reducing both types of absenteeism, suggesting that the package of EQUIP-T interventions helped motivate teachers to be in school, and in class, more of the time. The second is head teachers' use of performance appraisals as a method for teacher performance management. There is no evidence that EQUIP-T had an impact on teachers' participation in performance appraisals, which has remained at the same low level.

8.2 Conclusions on effectiveness, relevance, and sustainability

EQUIP-T's positive impact on pupil learning is highly commendable and is an important finding. It follows closely the interventions on early grade teacher training: the in-service training focused on Kiswahili in the first two years, and there was a positive impact on Kiswahili performance; the focus shifted to numeracy in the second two years, and then impact on maths performance was achieved. Given that participation in the in-service training is relatively high, and it has been broadly implemented as intended, it is likely that this aspect of the programme has made the biggest positive contribution to impact.

Furthermore, the generally low levels of pupil learning outcomes¹⁶⁵ highlight the relevance of EQUIP-T; its focus on in-service training and changing learning outcomes for children has been highly appropriate. However, when unpicking many of the intermediate outcomes expected to

¹⁶⁴ Conclusions about efficiency will be drawn in the final part of the endline based on the cost study.

¹⁶⁵ For example, only 18% of Standard 3 pupils are achieving the expected Standard 2 level in Kiswahili, 9% in maths.

contribute to pupil learning, according to the programme theory of change, the endline finds that the programme has not always been successful in achieving the expected effects, such as on teachers' and head teachers' practices. This raises the question of whether the potential impact on pupil learning could have been substantially larger, and of whether some linkages and assumptions in the theory of change proved to be valid.

Meanwhile, programme implementation has taken place during a period of external influences which are likely to have put negative pressure on the programme. The introduction of fee-free basic education in 2015 led to a surge in enrolment, with larger class sizes and more over-age pupils in the classroom. Staff turnover increased from the high levels already noted at midline, as the Government implemented a professionalisation policy, forcing out under-qualified head teachers and WEOs. In addition, delays in Government approval of materials and funds for districts were barriers to timely implementation, according to the EQUIP-T MA.¹⁶⁶ While the programme could have done more to achieve larger impact, it is recognised that it managed to achieve impact despite these factors.

This sub-section draws together key messages on the effectiveness of the various components measured in this part of the endline. Some of the areas of low effectiveness raise questions of relevance, and have implications for sustainability, and these issues are also discussed here.

8.2.1 Targeting improvements in pupil learning

Kiswahili performance improved most strongly between baseline and midline, reflecting the focus of the early grade teaching in-service training. The slowing of improvement after midline coincides with the more challenging context: class sizes grew substantially, the range in ages grew (increasing the proportion of over-age pupils), and more children came from poorer households and were occupied with more work outside the home.¹⁶⁷ At the same time, the maths results show that while these challenging contextual factors may have put downward pressure on maths performance, the focus on numeracy after midline in the in-service training played a part in mitigating this.

Improvements in girls' learning have been more pronounced than for boys in programme schools, with an attainment gap opening up in Kiswahili (with girls performing better than boys), and the boys' lead in maths narrowing. There is some evidence that EQUIP-T has contributed to larger gains for girls: teacher interactions with pupils became substantially more gender-balanced, though this improvement was seen between baseline and midline rather than after the specific gender response pedagogy training was rolled out after midline.

There has also been significant progress since baseline in narrowing the learning gap in literacy skills for children who do not speak Kiswahili (as a main language) at home, compared with those children who do. However, pupils from households who speak Kiswahili at home still perform far better than those who do not in both Kiswahili and maths, even after taking poverty into account. Since baseline, this impact evaluation has stressed that the overwhelming majority of pupils do not speak Kiswahili as their main language at home. As pointed out at midline, it is imperative for teacher training to take account of this barrier to learning and to support teachers with strategies to overcome it, in order to be relevant for the context. Whilst it is positive to see that the general teaching practices covered in the in-service training have supported a narrowing of the learning

¹⁶⁶ This information is from comments on the draft of this report from the EQUIP-T MA (received by the evaluation team on 15 November 2018).

¹⁶⁷ Whether the shift in pupil characteristics was due to a national change, or a rebalancing of the make-up due to the enrolment of pupils who previously would not have enrolled, is unknown.

gap in literacy skills, this raises the question of whether the programme should have been more tailored to address this issue.

8.2.2 Teachers' performance

At endline, almost all targeted teachers had attended the in-service training, although the coverage of modules was not complete. The incomplete coverage raises the serious challenge of a context with very high teacher turnover: just over half of early grade teachers from baseline are still in the same school at endline, although most moved within EQUIP-T districts. With such high turnover, as reported in the baseline and midline phases of the impact evaluation, and movement of teaching responsibilities between lower and upper grades, it is not surprising that many teachers have not attended training on all the modules.

The school-based in-service training model is intended to create a system of continuous learning and support such that all topics can be covered and sessions could be tailored to the topics identified by teachers. Schools' implementation of this training has been improving since 2015; however, there is still wide variation, with some schools holding far fewer sessions than expected, and targeted teachers appear to be attending less than half of the sessions. High turnover of teachers and INCOs is particularly problematic for this aspect of the training model. Challenges to attending school-based sessions reported by teachers include the lack of allowance, low motivation and inconvenient scheduling. Meanwhile, teachers want to see more training materials provided, while EQUIP-T is shifting away from a central model of training materials. These findings suggest a risk to the sustainability of the training model – although it is designed to be low-cost and locally-driven, capacity to maintain implementation appears to be low at the school level.

The endline finds mixed results on the use of positive teaching practices. Teachers' interaction with pupils across the classroom, a measure of inclusion of all pupils in the lesson, improved substantially between baseline and midline. However, this has not improved further since midline and in fact teachers are interacting less with pupils at the back of the room – likely a reflection of the larger class sizes. As mentioned above, the gender balance of interactions improved markedly between baseline and midline and this was sustained to endline, despite the sharp rise in classroom overcrowding. However, while teachers are demonstrating some positive behaviours more, use of many of the positive teaching practices measured by the survey has significantly worsened. Thus on the practices measured by this impact evaluation it is unclear if EQUIP-T has helped to improve teaching quality overall, although this may not be surprising in the context – particularly of such large class sizes. Teacher training needs to be tailored to ensure that teachers are able to deal with large pupil numbers.

Teachers are using different instructional materials more frequently in their lessons. However, despite distribution of TLMs to schools, EQUIP-T's materials are typically not found in classrooms and are not being used. This raises the question of whether the materials are relevant for teachers in these schools, since they are not choosing to use them, and whether the in-service training could have done more to support teachers in knowing how and why to use the materials.

Overall, there has been a reduction in teachers' absenteeism from the classroom since baseline, for Standards 1 and 2 teachers, and most of this drop happened before midline. As mentioned, there is weak evidence that EQUIP-T had an attributable impact in reducing overall teacher absenteeism from classroom, and levels are still extremely high. When absenteeism is factored into the timetabled weekly minutes, pupils lose around 40% of the official instructional time for both Kiswahili and maths. Teacher attendance in the classroom is a fundamental assumption underlying the programme theory of change, and yet absenteeism appears to be an endemic problem to

which the programme was only able to make a small improvement. This raises a question of whether the programme could have identified better ways to address this than its indirect route of improving motivation, and future programmes may need to introduce more comprehensive reforms if this indicator is to be shifted.

8.2.3 SLM

The programme has not been as effective as anticipated at rolling out SLM training: around three-quarters of head teachers have attended SLM training in the last two years, and those who have said the training was helpful in regard to knowing their responsibilities, and for teacher management. Again, the incomplete coverage is linked to high turnover, with only a quarter of the head teachers from baseline still in the school at endline. High turnover was highlighted in the midline evaluation, and this raises the question of whether the programme could have done more to mitigate the challenges of high turnover, in order to be a relevant and sustainable training model in this context.

The drop in the availability of SDPs since midline coincides with the tailing off of SLM training on SDPs, which largely took place in 2016, as well as the high turnover of head teachers, which means that a large group did not receive this training. This is unfortunate since the midline impact evaluation found that SDPs were seen as improving the relationship between head teachers and teachers, and schools and communities.

There are some positive improvements in head teacher leadership and management: head teachers are managing teachers' performance through more checking of lesson plans and tracking pupils' performance; and they are holding more staff meetings than at baseline and giving more rewards for teachers. On the other hand, the frequency of lesson observations by head teachers has fallen significantly since the midline.

EQUIP-T's SIS has so far not shown signs of effectiveness. Almost all schools have the tablets; however, less than half of head teachers have had any SIS training (again, likely a problem of turnover), and only a small proportion of schools have up-to-date enrolment and teacher numbers entered on the tablet, let alone any attendance data. There are a number of major problems which affect the relevance of this intervention: it takes too long to enter the data, most schools do not have functional electricity to make it easy to charge the tablet, and internet connectivity is poor. Given these challenges it is perhaps unsurprising that so few schools have entered the data and even fewer are using the SIS as a tool for school management. A more substantial assessment of needs and practicalities for entering and using the data during development might have helped to ensure this intervention was more relevant and thus more useful.

Over the four years since baseline the number of visits to schools by WEOs has increased significantly; qualitative research from midline suggests this was largely due to the provision of motorbikes and a WEO grant from EQUIP-T which can be used for travel costs. The programme plans to increasingly share the costs of the WEO grants with LGAs, and ultimately shift the cost entirely to LGA budgets (EQUIP-T MA, 2017). The success of this will be crucial for the sustainability of increased WEO visits and any support WEOs are expected to provide regularly at schools. However, despite the more frequent visits, at endline head teachers are less satisfied with WEOs and the support they receive from them than in the past, possibly connected to recent turnover in WEOs as part of the Government's professionalisation drive.

8.2.4 Community participation

More schools are holding annual meetings with parents and teachers than at baseline, a key communication and engagement mechanism. There has been an improvement in the number of SCs that have been trained on their roles and responsibilities. However, a sizeable minority of SCs are still not meeting regularly and head teachers do not feel the support from SCs is good. SCs have increased the focus on school finances in their meetings since baseline, perhaps reflecting the more regular receipt of capitation grants.

The PTP initiative has had some success: almost all schools have a PTP, and most have had training on their expected roles. However, PTPs are not very active in terms of meetings, and a sizeable minority did not take action to improve education in 2017. The activities that did take place were most commonly to improve pupil attendance and punctuality, although pupil attendance did not change between midline and endline and actually worsened slightly for boys. PTPs do not yet appear to be bridging the gap between schools and the wider parental body and community: most parents not in the PTPs do not know about them. Similarly, noticeboards were distributed to most but not all schools, yet only half of schools are displaying them, showing no improvement since baseline. Only a minority of parents are aware of the noticeboards and use them for information. Thus, these initiatives have not substantially improved the communication between schools and parents overall, and this raises the question of whether they are relevant to the context, or whether more effort is needed to make them effective.

EQUIP-T's grants to schools have been effectively administered: the targeted numbers of schools received the first PTP grant and IGA grant, largely in the correct amount, and most have spent the first PTP grant by endline. Most IGA grant recipient schools and communities have started implementing the IGAs, which is a promising sign for sustainability, as, if they are successful, in future the additional funds could be used to continue school improvement activities started under EQUIP-T.

8.2.5 Conducive learning environments for marginalised children

The impact evaluation confirms the importance of the programme's efforts to support marginalised or vulnerable children. A large minority of pupils come from households below the poverty line, most children did not eat before coming to school, almost a third had at least a 45-minute walk to get to school, and most children do not speak Kiswahili at home. Across various measures, pupils in the EQUIP-T districts are characterised by many dimensions of vulnerability and it is important that the programme as a whole recognises and responds in ways to overcome these barriers.

There has been relative success in rolling out the new conducive learning environments initiatives. Most schools have been trained on, and now have, JUU clubs, which are fairly active in meeting, and the majority carried out an activity in 2017. The focus of these activities was largely related to the school and learning environment, although some looked at pupil welfare and attendance. Most did not carry out activities to target specific groups of vulnerable pupils. Most schools have received the PTP girls' education grant, but only a quarter have so far spent it; the focus, as anticipated, has been on pupil welfare, infrastructure, and furniture, but more support may be needed to make sure all schools spend the grant. Most schools have received the *Shujaaz* magazine. Whilst it is harder to measure whether these outputs have led to the anticipated extra support for girls, children with disabilities and other marginalised children, the survey finds that schools have not yet integrated these issues into the SDPs, with very few SDPs containing activities to support girls' learning, pupil welfare, or vulnerable pupils. However, the new initiatives

targeting inclusion may have supported changes in teachers' classroom practices, contributing to the overall positive gains in girls' literacy and numeracy skills.

8.3 Recommendations

Recommendations have been split below between those for the EQUIP-T MA, and those for Government which are broader and more system-orientated. In relation to the EQUIP-T MA recommendations, it is recognised that at this stage of the programme many strategic decisions are taken jointly with senior government officials, and activities are largely planned and implemented by local government, supported by EQUIP-T staff. This means that most recommendations require co-ordination and action from both parties.

8.3.1 Recommendations for EQUIP-T MA

These recommendations are intended for EQUIP-T MA. Recognising that the programme will close in January 2020, the recommendations focus largely on reinforcing the existing interventions, rather than introducing new initiatives.

1. Given the high level of teacher and head teacher turnover and the findings on incomplete attendance at training and module coverage, the programme will struggle to achieve the full intended impact of the trainings. The programme should consider running refresher training or mop-up training, in order to ensure all serving teachers, head teachers and WEOs have received the intended training. This particularly relates to:
 - a. the early grade in-service training, in relation to which the proportion of targeted teachers who have not covered all the modules is largest for Kiswahili, followed by numeracy, and finally gender-responsive pedagogy;
 - b. SLM training, in relation to which almost one-third of head teachers have never received any training on SDPs, and more than half have not received training on the SIS. This links to the poor results in terms of availability and implementation of SDPs and use of the SIS; and
 - c. refresher training for WEOs across various components: SLM, SCs, and PTPs; although this survey did not collect information on training for WEOs, the high turnover of WEOs suggests that such refresher training would also be valuable.
2. Similarly, as increased and strengthened engagement of communities and parents in education is considered critical in sustaining the capacity and activity of SCs and PTPs after EQUIP-T ends, it is important to try to also reach the untrained groups of SCs and PTPs with the training. This would help to ensure that all schools spend the second PTP grant.
3. Where in-service training is being offered, EQUIP-T, together with its local government partners, should consider ways to overcome some of the challenges identified by participants. Issues such as the location of the venue, timing and duration of sessions, volume of training materials, full coverage of out-of-pocket expenses (and potentially additional incentives or compensation—not necessarily monetary), if addressed could help to ensure better attendance.
4. As part of refresher training for teachers, EQUIP-T, together with TIE, should consider putting emphasis on using the TLMs, particularly reading materials, distributed by the programme, to increase their use.
5. Recognising the contextual challenges highlighted across the stages of this impact evaluation, the programme should support the Government by either adapting existing modules or developing new modules to focus on managing varied class sizes and the issue of pupils' home language. The programme's forthcoming general pedagogy module would be a good opportunity to include this material.

6. The programme should work with PO-RALG to review and reduce the templates and information required to be entered in the SIS. At the moment, entering the data takes too long, and some head teachers do not understand how to carry out the data entry, and as a result the SIS is barely being used at all. The issue could be that too much data are expected to be entered, or the issue may relate to the functionality and simplicity of the software package: both should be looked at. The programme should also investigate why tablets are not functioning properly, and provide a maintenance system to allow tablets to be quickly repaired.
7. EQUIP-T should find ways to support its local government partners to strengthen the role of the PTP in building relationships between the school and wider community, and use refresher training to emphasise this.

8.3.2 Recommendations for Government

With EQUIP-T coming to an end in January 2020, the following recommendations are identified for the Government of Tanzania in deciding how to take forward the learning from the EQUIP-T programme. DFID may also reflect on these recommendations in considering how it will continue to support the Government after EQUIP-T closes. At this stage, these are interim recommendations which will be finalised following the completion of the qualitative research and costing parts of the endline.

1. A number of contextual factors are found to place limits on the extent of success of the EQUIP-T intervention. The Government should continue its efforts to minimise these challenges, for example:
 - a. addressing factors that contribute to staff turnover beyond retirement and other inherent reasons for attrition—this could include emphasis on continued professional development that is linked to some form of professional reward;
 - b. responding with additional teachers and infrastructure to the surge in enrolment, which has led to larger class sizes and more double shifts;
 - c. having an explicit policy and set of tools for how teachers should best deal with the language barriers for children who do not speak Kiswahili at home in primary classrooms—this would be beneficial in guiding in-service materials and future training for teachers.
2. If the Government is adopting elements of the EQUIP-T in-service training for early grade teachers into a national in-service training programme, it should also consider ways to encourage greater attendance of teachers, particularly in the school-level component, to ensure coverage of modules. It should also ensure the training content supports teachers in dealing with large class sizes and pupils with different home languages.
3. Government, through TIE, should put greater emphasis on the use of relevant TLM in early grade classroom teaching, as part of the content of future national in-service training programmes.
4. The Government should focus on improving teacher school and classroom attendance, which has a large impact on the number of instructional hours pupils receive. Teacher motivation is one component of attendance, and low motivation has multiple causes. However, another factor is the accountability mechanism, including management and monitoring by head teachers and WEOs. Strategies for monitoring and managing teachers' classroom absenteeism should be a key part of the teacher management element of SLM training, spearheaded by ADEM.
5. Reducing pupil absenteeism could usefully be another focal area for head teachers under the ADEM-led SLM training, recognising that timely information on pupil attendance patterns (potentially via a functioning SIS) and regular engagement with parents to understand barriers to attendance are essential to addressing this.

6. When considering the role of SCs, the impact evaluation finds that even when SCs have received training, this is not enough to ensure they actively engage with school matters. Government should consider ways to encourage SCs to be more active in their meetings and the support they provide to the school.
7. Parents do not typically know about the PTPs or the school noticeboards, and those who do know about noticeboards do not check the information posted on them. Taken together, these findings strongly imply that using PTPs and noticeboards to improve information sharing and relationships between the school and the wider parental body is unlikely to be effective as things stand (although PTPs may help meet classroom support objectives). The Government should thus be cautious if it is considering rolling out these initiatives nationwide to meet these broader objectives.

8.4 Lessons learned

Many of the findings above, and the recommendations, may be useful for a wider audience when considering the experience of implementing a large education programme. This section draws out some specific lessons which could help inform the design of future programmes in similar contexts to the EQUIP-T regions of Tanzania. These lessons are likely to be further refined on the basis of the qualitative research and cost analysis in the second part of the endline evaluation.

1. In a context where children face multiple aspects of vulnerability and marginalisation, programmes should look at how they can tackle various barriers in a holistic package. For example, improving the quality of teaching in the classroom will have limited potential to increase pupils' learning if pupils are coming to school tired from additional work and hungry due to lack of food. These factors, though outside of the education system, are likely to affect children's ability to concentrate and learn, and thus are an important part of ensuring access and learning.
2. Programmes which rely on a substantial component of in-service training, whether for teachers, head teachers, or their supervisors, need to pay attention to a context of high turnover. The training should build in revision or catch-up sessions, to allow all targeted beneficiaries to access the training and intermittently receive a refresher to consolidate the learning. Where possible, strategies to reduce turnover would complement the programme. In addition, programmes need to consider elements to ensure high attendance and to minimise impact on day-to-day responsibilities: avoid holding teacher training during teaching time, and ensure incentives are aligned to ensure participants attend training (including avoiding leaving participants out of pocket). Many of these lessons – refresher training, scheduling, and compensation – also apply to training for community members, such as SCs.
3. The content of teacher training needs to be tailored to the realities of the classroom, which could include managing large class sizes and the divergence between language of instruction and pupils' home language. Training which does not acknowledge these constraints and actual classroom context can leave teachers with skills they are not able to put to use.
4. A number of education programmes have been found to have successfully distributed TLMs, and yet those materials were not actually in use in classrooms (Glewwe and Muralidharan, 2015). When designing new programmes, effort should go into ensuring that any new materials are developed based on what teachers will find useful and practical, and the roll-out should have strategies to support teachers with using the materials in their lessons.
5. Interventions to improve data collection, management, and use at the school level and above, such as EQUIP-T's SIS, should look at the practicalities of entering the data (from software and hardware perspectives), and what data will really be used, and should start with something very simple to get the process off the ground. The hardware needs to be able to function properly, with adequate electricity and internet, and there needs to be a plan for maintenance if repairs are needed.

8.5 Implications for final phase of the impact evaluation

This initial phase of the endline evaluation has found a number of mixed results, with some components effectively rolled out but not resulting in the changes that were anticipated. The final part of the impact evaluation endline provides the opportunity for the evaluation team to use qualitative research to better understand the reasons for some of the key findings in this report. An initial options paper for the design of the endline qualitative study was shared with DFID in December 2018. This proposed two alternatives:

1. An exploratory approach to understand programme impact, based on triangulating the perceptions of national and district stakeholders closest to the programme, including the EQUIP-T MA, DEOs, WEOs, with observations and discussions in a small purposively selected sample of schools. The method is inductive in that it seeks to explore or generate fresh reasons to explain impact.
2. An explanatory approach to examine programme impact, which would build on the hypothesis put forward in this report that the in-service training is likely to have been the strongest contributor to the positive impact seen on learning outcomes. This method is deductive, and it would investigate how the in-service training has been understood, valued and used in classrooms, and whether stakeholders in a small purposive sample of schools perceive the in-service training's influence on improved teacher performance. An additional focus of the study would be on community engagement, particularly the PTP initiative, to try to understand from school-level stakeholders whether and how the PTP has worked to support the class level and to delve deeper into why it may not have been successful in strengthening communication between the wider parental body and the school. This secondary focus is proposed for two main reasons. First, the quantitative evaluation results related to community engagement are limited in scope because the type of changes anticipated in the programme theory of change are difficult to measure quantitatively. Second, considering the strong views of the EQUIP-T MA that PTPs are supporting schools more effectively than the quantitative data suggests in this report.¹⁶⁸

The qualitative endline also provides an opportunity to evaluate component 3, strengthened district planning and management, as this is the only original component which has been almost entirely excluded from the quantitative endline.

The design of the endline qualitative study is currently being refined, following feedback from DFID on the options paper. A full draft design will be set out in the Draft Endline Planning Report Part II, which will be circulated to the study's Reference Group in late February 2019 for feedback. This planning report will also contain a proposal for the design of the endline cost study.

¹⁶⁸ In the EQUIP-T MA feedback on the draft of this report (received by the evaluation team on 15 November 2018), various specific examples are cited of successful PTP actions. One of the comments is 'Many people are of the opinion that PTPs have contributed more than is argued in this report.'. The same type of comment was not made about any of the other EQUIP-T interventions.

References

- Abadzi, H. (2006) 'Efficient learning for the poor: Insights from the frontier of cognitive neuroscience'. World Bank, Directions in Development Series.
- Bennell, P. and Akyeampong, K. (2007) 'Teacher Motivation in Sub-Saharan Africa and South Asia'. *Researching the Issues*, 71, London, UK DFID.
- Blundell, R. and Costa Dias, M. (2000) 'Evaluation Methods for Non-Experimental Data'. *Fiscal Studies* 21, no. 4: 427–68. www.jstor.org/stable/24437670.
- Boyden, J. and Ennew, J. (eds.)/Save the Children (1997) *Children in Focus – a Manual for Participatory Research with Children*. Stockholm: Save the Children Sweden
- Caliendo, M. and Kopeinig, S. (2005) 'Some practical guidance for the implementation of propensity score matching'. *IZA Discussion Papers*, No. 1588.
- Cambridge Education (2014) 'Final EQUIP-Tanzania Inception Report', 5 February.
- Cartwright, N. and Hardie, J. (2012) *Evidence Based Policy: A Practical Guide To Doing It Better*. Oxford: Oxford University Press.
- Carvalho, S. and White, H. (1997) 'Combining The Quantitative and Qualitative Approaches to Poverty Measurement and Analysis', Technical Paper 366. Washington DC: World Bank.
- Cueto, S. *et al.* (2009) 'Psychometric characteristics of cognitive development and achievements instruments in Round 2 of Young Lives'. Young Lives Technical Note #15, January 2009.
- De Grauwe, A. (2001) *School Supervision in Four African Countries: Vol. I: Challenges and Reforms*. Paris: UNESCO IIEP.
- De Ree, J. (2016) 'How Much Teachers Know and How Much it Matters in Class: Analyzing Three Rounds of Subject-Specific Test Score Data of Indonesian Students and Teachers'. World Bank Policy Research Working Paper
- Diseko, E., Barkhuizen, N. and Schutte, N. (2015) 'The Relationship between Talent Management and Turnover Intentions of Teachers in Botswana'. 20th International Academic Conference
- DFID (2013) 'Intervention Summary, Education Quality Improvement Programme in Tanzania (EQUIP-T)'. London, UK DFID.
- DFID (2011) *DFID Ethics Principles for Research and Evaluation*.
- Dowd, A. *et al.* (2013) 'Literacy Boost Cross Country Analysis Results', Save the Children, Washington D.C.
- Drake, L., Woolnough, A., Burbano, C. and Bundy, D. (2016) *Global school feeding sourcebook: lessons from 14 countries*. London, UK: Imperial College Press.
- EQUIP-T MA (2015) 'EQUIP-Tanzania Annual Report 2015'. United Republic of Tanzania (URT), PMO-RALG, MOEVT, and DFID.
- EQUIP-T MA (2016) 'EQUIP Tanzania, 2016 Annual Report External Version'. URT, DFID.
- EQUIP-T MA (2017) 'EQUIP Tanzania, 2016-17 Annual Summary External Version'. URT, DFID.
- Garbarino, S., and Holland, J. (2009) 'Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results', Issues paper commissioned by DFID. GSDRC Emerging Issues Service.
- Gershoff, E.T. (2017) 'School corporal punishment in global perspective: prevalence, outcomes, and efforts at intervention'. *Psychology, Health & Medicine* 22, 224–239. <https://doi.org/10.1080/13548506.2016.1271955>

- Glewwe, P and Muralidharan, K. (2015) 'Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps and Policy Implications'. RISE-WP-15/001. RISE Programme.
- Glewwe, P., Kremer, M. and Moulin, S. (2009). 'Many children left behind? Textbooks and test scores in Kenya'. *American Economic Journal: Applied Economics* 1 (1) (January 2009): 112-35
- Gove, A., Brunnet, T., Bulat, J., Carrol, B., Henny C., Macon, W., Nderu E., and Sitabkan Y. (2017) 'Assessing the impact of early learning programs in Africa.' In Kenneth R Pugh, Peggy McCardle & Annie Stutzman (Eds) *Global Approaches to Early Learning Research and Practice*. New Directions for Child and Adolescent Development. 158, 25-41.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). 'Toward a conceptual framework for mixed-method evaluation designs'. *Educational evaluation and policy analysis*, 11(3), 255-274.
- Guest, G., MacQueen, K. and Namey, E. (2012) *Applied Thematic Analysis*. Sage Publications.
- Hallinger, P. and Heck, R. (1996) 'The Principal's Role in School Effectiveness: An Assessment of Methodological Progress, 1980-1995'. In K. Leithwood, *The International Handbook of Research on Educational Leadership and Administration*. New York: Kluwer Press.
- Hardman, F. and Dachi, H. (2012) 'Evaluation of School-Based INSET Pilot Programme'. York, UK: Institute for Effective Education, University of York.
- Hattie (2009) 'Visible learning: A synthesis of over 800 meta-analyses related to achievement.' London; New York, Routledge.
- Hoyland, A. Dye, L and Mawton, C (2009) 'A systematic review of the effect of breakfast on the cognitive performance of children and adolescents'. *Nutrition Research Reviews* 22.2: 220-243.
- Imbens, G. and Rubin, D. (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.
- J-PAL (2014) 'Increasing test score performance: What interventions are most effective at increasing student learning.'
- Krippendorff, K. (2004) *Content Analysis: An Introduction to Its Methodology*. Sage Publications.
- Lechner, M. (2002) 'Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies'. *Review of Economics and Statistics*, 84(2), 205-220.
- Mays, N. and Pope, C. (1995) 'Rigour and Qualitative Research', *British Medical Journal* 311(6997): 109-112.
- Metzler, J. and Woessmann, L. (2012) 'The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation'. *Journal of Development Economics*, 99 (2)
- MOEST (2018) *Education Sector Development Plan 2016-17 to 2020-21 Tanzania Mainland*. URT
- MOEST (2017) *National Strategy on Inclusive Education*. URT, MOEST.
- MOEST and PO-RALG (2018a) *Tanzania Education Program for Results (EPforR) Programme Operation Manual*. URT, GOT.
- MOEST and PO-RALG (2018b) *Tanzania Education Program for Results (EPforR) Annual Report 2017-18 Draft (09 October 2018)*. URT, GOT.
- MOEVT (2016) *Curriculum for Basic Education Standards I and II*. URT, MOEVT.
- MOEVT (2009a) *Basic standards for pre- and primary education in Tanzania*. Dar es Salaam: MOEVT. www.ed-dpg.or.tz/pdf/PE/Basic%20Standards%20for%20Pre%20and%20Primary%20Education%20in%20TZ_2009.pdf

- MOEVT (2005a) *Mathematics Syllabus for Primary Schools Standard I-VII*. URT, MOEVT.
- MOEVT (2005b) *Kiswahili Syllabus for Primary Schools Standard I-VII*. URT, MOEVT.
- Mulkeen, A. (2010) 'Teachers in Anglophone Africa. Issues in Teacher Supply, Training and Management'. Washington, DC: The World Bank.
- Nag, S., Chiat, S., Torgerson, C. and Snowling, M.J. (2014) 'Literacy, Foundation Learning and Assessment in Developing Countries: Final Report'. Education Rigorous Literature Review. DFID.
- Ogando Portela, M.J. and Pells, K. (2015) 'Corporal Punishment in Schools – Longitudinal Evidence from Ethiopia, India, Peru and Vietnam'. Florence: Innocenti Discussion Papers.
- OPM (2015a) EQUIP-Tanzania Impact Evaluation Final Baseline Technical Report, Volume I: Results and Discussion. OPM.
- OPM (2015b) EQUIP-Tanzania Impact Evaluation Final Baseline Technical Report, Volume II: Methods and Technical Annexes. OPM.
- OPM (2016a) 'EQUIP-Tanzania Impact Evaluation: Midline Planning Report'. OPM.
- OPM (2016b) 'EQUIP-Tanzania Impact Evaluation: Preliminary Indicators for the Programme Treatment Districts: Results from the Baseline and Midline Quantitative Surveys'. OPM.
- OPM (2016c) 'EQUIP-Tanzania Impact Evaluation: Propensity Score Analysis Review: A Technical Note'. OPM.
- OPM (2016d) 'EQUIP-Tanzania Impact Evaluation: Midline Quantitative Fieldwork Report'. OPM.
- OPM (2017a) 'EQUIP-Tanzania Impact Evaluation Final Midline Technical Report, Volume I: Results and Discussion'. OPM.
- OPM (2017b) 'EQUIP-Tanzania Impact Evaluation Final Midline Technical Report, Volume II: Methods and Supplementary Evidence'. OPM.
- OPM (2018) 'EQUIP-Tanzania Impact Evaluation: Endline Planning Report: Part I Quantitative Research'. OPM.
- Osim, R.O., Chika, C. and Uchendu, I. O. (2012) 'Class size pressure: An Impediment to teacher's work'. *Global Advanced Research Journal of Educational Research and Review* Vol 1(5)
- Ouane, A. and Glanz, C (2011) 'Executive Summary'. In Ouane, A and Glanz, C (eds.) *Optimizing Learning, Education and Publishing in Africa: the Language Factor. A Review and Analysis of Theory and Practice in Mother-Tongue and Bilingual Education in sub-Saharan Africa*. UNESCO Institute for Lifelong Learning and the Association for the Development of Education in Africa / African Development Bank
- Paternoster, R., Brame, R., Mazerolle, P. and Piquero, A. (1998) 'Using the correct statistical test for the equality of regression coefficients'. *Criminology*, 36: 859–866. doi:10.1111/j.1745-9125.1998.tb01268.
- Pivik, R.T. et al. (2012) 'Eating breakfast enhances the efficiency of neural networks engaged during mental arithmetic in school-aged children' *Physiology and Behaviour* 106.4 2012: 548-555.
- 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.

- Rosenbaum, P.R. and Rubin, D.B. (1985) 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score'. *The American Statistician* 1985: 39:33–38.
- RTI (2016) 'Assistance to Basic Education All Children Reading (ABE ACR): Preliminary Findings Report, Tanzania National Early Grade Reading Assessment (EGRA)'. North Carolina. USAID.
- RTI (2014) 'National Baseline Assessment for the 3Rs (Reading, Writing, and Arithmetic). Using EGRA, EGMA, and SSME in Tanzania'. Study Report. Draft. Washington, DC: USAID.
- Robinson, Lloyd and Rowe (2008) 'The Impact of Leadership in Student Outcomes: An Analysis of the Differential Effects of Leadership Types', *Education Administration Quarterly*.
- Rubin, D. (2001) 'Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation', *Health Services & Outcomes Research Methodology* 2, 169–188.
- Sabarwal, S., Evans, D. and Marshak, A. (2014) 'The permanent input hypothesis: the case of textbooks and (no) student learning in Sierra Leone'. Policy Research working paper, no. WPS 7021. Washington, DC: World Bank Group
- Schuh Moore, A.-M., DeStefano, J., and Adelman, E. (2012) 'Opportunity to Learn: A high impact strategy for improving educational outcomes in developing countries'. Washington, DC: USAID. Available at: <https://www.fhi360.org/resource/opportunity-learn-high-impact-strategy-improving-educational-outcomes-developing-countries>
- Singh, A. (2015a) <https://blogs.worldbank.org/impactevaluations/how-standard-standard-deviation-cautionary-note-using-sds-compare-across-impact-evaluations>
- Singh, A. (2015b) 'Private school effects in urban and rural India: Panel estimates at primary and secondary school ages.' *Journal of Development Economics*, Volume 113, March 2015 pp16-32.
- Siraj, I., Taggart, B., Melhuish, E., Sammons, P. and Sylva, K. (2014) *Exploring Effective Pedagogy in Primary Schools: Evidence from Research*. London: Pearson. Available at: https://research.pearson.com/content/plc/prkc/uk/open-ideas/en/articles/explore-eppse/jcr_content/par/articledownloadcompo/file.res/Exploring%20Effective%20Pedagogy%20in%20Primary%20Schools.pdf
- The National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research (1979) 'The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Department of Health, Education, and Welfare'. [www.hhs.gov/ohrp/humansubjects/guidance/belmont.html]
- The Open University (no date) 'Ethics Principles for Research Involving Human Participants', [www.open.ac.uk/research/main/sites/www.open.ac.uk.research.main/files/files/OU%20Ethics%20Principles%20for%20Research%20Involving%20Human%20Participant.pdf]
- UN Committee on the Rights of the Child (CRC) (2013) Consideration of reports submitted by States parties under article 44 of the Convention, Third to fifth periodic reports of States parties due in 2012 : United Republic of Tanzania, 4 November 2013, CRC/C/TZA/3-5. Available at: <http://www.refworld.org/docid/54ae89254.html> [accessed 26 October 2018]
- UNESCO (2016) 'If you don't understand how can you learn?' GEM Report Policy Paper 24. Paris: UNESCO.
- UNESCO (2015) 'Education For All Global Monitoring Report 2015: Education for All 2000–2015: Achievements and Challenges'. Paris: UNESCO.
- UNESCO (2014) 'Education For All Global Monitoring Report 2013/14: Teaching and Learning: Achieving Quality for All'. Paris: UNESCO
- UNESCO (2007) 'Education For All Global Monitoring Report 2008: Education for All: Will we make it?' Paris: UNESCO.

- UNESCO (2005) *Guidelines for Inclusion: Ensuring Access to Education for All*. UNESCO
- UNICEF (2009) *Child Friendly Schools Manual*. New York
- Uwezo (2011) 'Are our Children Learning? Annual Learning Assessment Report 2011'. Dar es Salaam: Uwezo Tanzania.
- US Department of Health and Human Services (2009) 'Code of Federal Regulations TITLE 45 Public Welfare. PART 46. Protection of human subjects'.
[www.hhs.gov/ohrp/policy/ohrpreulations.pdf]
- Vogel, I. (2012) 'Review of the use of theory of change in international development'. Review paper for the Department of International Development. DFID.
- Westbrook, J., Durrani, N., Brown, R., Orr, D., Pryor, J., Boddy, J. and Salvi, F. (2013) 'Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries'. Final Report. Education Rigorous Literature Review. UK DFID. Available at:
www.gov.uk/government/uploads/system/uploads/attachment_data/file/305154/Pedagogy-curriculum-teaching-practices-education.pdf
- White, H. (2009) 'Theory-Based Impact Evaluation: Principles and Practice'. 3ie Working Paper 3. International Initiative on Impact Evaluation.
- Wright, B and Stone, M. (1979) *Best Test Design*. Chicago: Mesa Press.

Annex A Agreed terms of reference

During the contracting and inception phases it was agreed by DFID and OPM that the scope of the impact evaluation needed to be reduced from that outlined in the original terms of reference (TOR) and that not all of the original TOR objectives could be met by the impact evaluation.¹⁶⁹

This section begins by setting out the original impact evaluation purpose and then discusses the implications of DFID's design choices during the contracting and inception phases. It goes on to outline what the impact evaluation will measure and the evaluation questions to be answered, and then sets out the revised impact evaluation purpose.

A.1 Impact evaluation purpose in original TOR

According to the original TOR the purpose of the impact evaluation of the EQUIP-T programme is twofold:

- “Assess if the EQUIP-T interventions in supported councils [districts] contribute to better basic learning outcomes amongst primary school age students.”; and
- Assess “which specific support interventions and measures of quality service provision were most significant in improving learning outcomes and to what extent are these replicable and affordable in the Tanzanian/E. African context.”

The original TOR also specified that:

- “The IE must ensure that the evidence is used to promote lesson learning, accountability, and understanding of the cost effectiveness and potential of the programme's intervention and approach.”

A.2 Impact evaluation design options

OPM's technical proposal (May 2013) provided DFID three impact evaluation design approaches to choose between.¹⁷⁰

The first option, the gold standard randomised control trial (RCT) approach would have been able to identify which specific EQUIP-T interventions were most effective in improving learning outcomes and programme scalability, but would have required the specification of multiple treatment groups across which to randomise assignment of programme exposure.

The second option, the hybrid approach, offered an intermediate option. Under this approach, a base package of EQUIP-T interventions would have been implemented in all treatment schools. Then the treatment group of schools would have been split into treatment sub-groups with an additional EQUIP-T programme intervention randomly assigned to each sub-group. This would have allowed for the assessment of whether adding specific EQUIP-T interventions to the base package led to further relative improvements in the key indicators and scalability.

For the final option, the basic approach, no attempt would be made to control the roll-out of specific EQUIP-T interventions within the EQUIP-T programme districts, allowing assessment of the impact

¹⁶⁹ DFID-OPM correspondence March 10, 2014.

¹⁷⁰ OPM's technical proposal is available on request.

of the EQUIP-T programme as a whole, but not of relative importance of different EQUIP-T components in improving learning outcomes or scalability.

The technical proposal also provided two options for the sample size for the quantitative baseline survey. The first option was for 100 EQUIP-T (treatment) schools and 100 non-EQUIP-T (control) schools, a total of 200 primary schools. The changes in the baseline proportion of pupils meeting Kiswahili and mathematics proficiency requirements detectable with this sample size are given in the EQUIP-T Impact Evaluation Inception Report (OPM 2014a). This was the minimum sample size to detect the expected effect size changes. A second option that would yield higher precision and improve the ability to detect EQUIP-T programme effects was offered, with a sample of 150 treatment schools and 150 control schools, a total of 300 primary schools.

For all three quantitative design approaches outlined above and provided in the technical proposal, the qualitative research would provide additional detail on issues around gender, reasons for observed changes in pupil learning levels, data on district education management and community participation in and demand for accountability in education. These qualitative data would be complementary to the quantitative survey data, but would not in themselves provide a theory-based evaluation or a rigorous attribution of impact to different EQUIP-T components.

A.3 DFID design choices

The three quantitative design options: gold standard, hybrid and basic approaches (see section A.2), were discussed during the contracting and inception phases and DFID selected the basic approach because of a preference for implementation of all EQUIP-T programme components in all EQUIP-T districts at approximately the same time and for cost considerations, as the gold standard and hybrid approaches would have required larger sample sizes and additional research activities and therefore would have been more costly than the basic approach.

The two sample size options provided: a total of 200 vs 300 schools were also discussed during contracting and inception and DFID selected the 200 school sample size.

Based on discussions with DFID and comments from the Specialist Evaluation and Quality Assurance Services (SEQAS), the qualitative research design was revised to include development of **an enhanced programme TOC** including contextual factors, priority parts of which will be tested during the follow-up rounds of the qualitative fieldwork.

A.4 What the impact evaluation will measure under the agreed terms of reference

The main focus of the impact evaluation will be to measure any EQUIP-T impact in the EQUIP-T programme districts covered by the IE and to provide accountability for the UK taxpayer in terms of the impact of resources used. The impact evaluation will also provide evidence on programme cost-effectiveness and fiscal affordability (separate fiscal study), promote lesson learning across districts and provide indications to DFID and the Government on which EQUIP-T programme components may likely be more effective in improving pupil learning outcomes.

Original TOR purpose 1

The evaluation will provide quantitative evidence on the impact of the EQUIP-T programme on learning outcomes for primary school pupils supported by qualitative research findings that will probe gender aspects and reasons for changes in pupil learning levels.

Original TOR purpose 2

Given DFID's choice of the basic approach (section A.3), the impact evaluation will explore other possibilities for understanding which EQUIP-T components may be more effective in improving pupil learning outcomes (assuming there is impact) and scalability.

The quantitative component will explore whether it will be possible to take advantage of any naturally occurring variation in roll-out of specific EQUIP-T interventions within the evaluation treatment sample, in order to identify impact of specific interventions. However, without random assignment of specific interventions or without stratifying the sampling of treatment schools by package of interventions (see above) it is unlikely that there would be enough variation in the sample to robustly identify differential impact.

It should be noted that the original TOR did not specify a theory-based impact evaluation nor was the development of a TOC required beyond that developed by the MA part of the original TOR for the IE. However, in light of SEQAS comments and discussions with DFID, the IE design has been revised to set out a process whereby the qualitative research will develop an enhanced TOC including contextual factors.

The EQUIP-T programme TOC will inform the IE as a whole, but is particularly important for the qualitative component because it should permit stronger **generalisation and some attribution of impact**. Specifically, the EQUIP-T TOC change will be used to map out EQUIP-T's causal chain and the contextual assumptions that must hold for EQUIP-T activities to lead to the desired impact (following the approach set out in White 2009). The IE will use (primarily) qualitative data to conduct 'rigorous factual analysis' on whether the expected links in the causal chains hold and whether the assumptions are valid over time, for some of the links in the causal chain selected on the basis of their perceived importance by key stakeholders.

While this is not a theory-based evaluation in the pure sense because it is not comprehensive on all causal pathways, the IE will use theory to produce results on which components of EQUIP-T are likely to contribute to changes in key outcomes and outputs in different contexts. This will yield **indicative results on which interventions were perceived to be more effective**, and coupled with secondary data analysis of the context in other areas of Tanzania to check whether these contextual assumptions hold there as well, this will enable **consideration of the likely impact of EQUIP-T if implemented at scale**.

Following the discussions and agreements with DFID during the contracting and inception phases, the primary aim of the impact evaluation will be to measure the impact of EQUIP-T over time. To do this the design of the quantitative component seeks to maximise internal validity. The EQUIP-T regions and districts were purposively selected by the MA on the basis of region rankings and district rankings in terms of education performance and financial resources and include primarily rural districts (see OPM 2014a). A large majority of rural districts in Tanzania share similar characteristics and therefore although the IE impact results will not be statistically generalizable outside the IE sample, it is **reasonable to expect that the findings will have some applicability in other districts as well if sufficiently similar to the treatment districts**, other things being equal. The impact evaluation will use, among other things, the rich dataset compiled for the quantitative baseline sampling frame to compare EQUIP-T districts along several key characteristics including education performance, infrastructure, poverty measures and population density, to similar districts not participating in EQUIP-T to assess the potential for generalisation.

The impact evaluation will also assess cost-effectiveness of the EQUIP-T programme and the fiscal affordability of rolling out EQUIP-T to regions and districts beyond the initial programme areas in a separate fiscal study (see OPM 2014a).

A.5 Evaluation questions

The original TOR specified key questions related to the OECD-DAC evaluation themes of relevance, effectiveness, impact and sustainability, to be answered by the impact evaluation. These questions are shown in Table 20, together with what the impact evaluation will measure, given the EQUIP-T programme design and changes to the TOR agreed with DFID.

Table 20 What the impact evaluation will measure

Original terms of reference evaluation questions	Measured by the impact evaluation under agreed TOR
“Have the programme interventions targeted the most necessary, most economical and appropriate combination of interventions for improvements in the quality of education?”	Partly, as the EQUIP-T components had already been determined by the MA during their inception phase and differential roll out of different EQUIP-T components was not deemed possible by DFID. The qualitative component will examine which EQUIP-T components were perceived to raise education quality.
“Has pupil-teacher ‘time on task’ been significantly increased in target schools?”	Yes.
“Are better pedagogic practices that promote effective learning, demonstrably in place?”	Yes.
“Have the target councils been able to increase learning outcomes for girls / boys, including disadvantaged children, beyond those more generally obtained in comparable areas?”	Yes.
“Do councils have costed plans in place that are realistic both fiscally and institutionally for the long term maintenance of quality within schools including provision and quality of teachers, operations, inputs and maintenance of school infrastructure?”	Partly if possible. The qualitative component will through the district level interviews attempt to collect information on the availability of costed plans, but not their quality, for the EQUIP-T programme councils (districts) selected as qualitative research sites.
“Improved education quality.”	Yes.
“Improved teaching of early-grade reading and numeracy resulting in more children able to read with comprehension” and with curriculum appropriate numeracy skills.”	Yes.
“Improved teaching of early-grade reading and numeracy resulting in more children able to read with comprehension” and with curriculum appropriate numeracy skills.”	Yes.
“More time on task for primary school children, resulting in more children passing their end of primary school examinations”	Yes.
“More girls able to make the transition to secondary school”.	No, as the EQUIP-T programme will focus on the early grades and impact of the programme, if any, on this outcome would be highly unlikely to be detectable within the life of the impact evaluation.
Pupil learning results should be disaggregated by gender.	Yes.

The impact evaluation should examine impact for disabled children.	No, because the EQUIP-T programme does not contain any component or activities aimed at this particular group ¹⁷¹ .
The impact evaluation should include poverty measures for pupils.	Yes.

A.6 Revised purpose of the impact evaluation

Thus the impact evaluation will:

- Generate evidence on impact of EQUIP-T on learning outcomes for pupils in primary education, including any differential impacts for girls and boys;
- Provide evidence on the fiscal affordability of scaling up EQUIP-T beyond the initial EQUIP-T regions and districts (separate fiscal study);
- Assess perceptions of effectiveness of different EQUIP-T components through the qualitative research and explore possibilities to do so through the quantitative component; and
- Communicate evidence generated by the impact evaluation to policymakers and key education stakeholders, including DFID and MOEVT¹⁷² to promote accountability and lesson learning.

A.7 Changes to the impact evaluation design since the technical proposal

In addition to the reduction of scope of the TOR outlined above the following changes to the IE design compared to the technical proposal were made based on discussions with DFID during the inception phase and feedback from the first Reference Group meeting for the impact evaluation¹⁷³.

- Scope of impact evaluation expanded for the qualitative component to examine perceived EQUIP-T contributions to changes in relevant outcomes and outputs, to the extent possible within the scope of the IE;
- Quantitative fieldwork to start in March 2014 (first start date was October 2013, second start date was January 2014);
- Use of EGRA/EGMA style pupil learning assessments instead of UWEZO assessment testing 3,000 standard 3 pupils in Kiswahili (EGRA) and mathematics (EGMA);
- Test standard 3 pupils (Kiswahili and mathematics) instead of standard 2 and standard 5 pupils;
- Administer teacher development needs assessment (TDNA) to standards 1-3 (Kiswahili and mathematics) and 4-7 (mathematics) teachers instead of to standard 2 and standard 5 teachers;
- One standard 2 Kiswahili and one standard 2 mathematics lesson will be observed for each sample school instead of one standard 3 lesson and one standard 5 lesson;
- Replace pupil tracer survey to collect data for poverty measure by data collected at school level (from tested pupils' parents);

¹⁷¹ MA-OPM and DFID-OPM correspondence December 2, 2013.

¹⁷² Prior to the new government which took office in November 2015, the Ministry had responsibility for vocational training and was known as the Ministry of Education and Vocational training (MOEVT).

¹⁷³ The draft inception report was subsequently further revised based on comments received from the SEQAS review on March 7, 2014.

- To obtain school sample: in the second stage, match control schools to treatment schools using PSM instead of random selection;
- 17 EQUIP-T programme (treatment) districts in the five EQUIP-T regions covered by the impact evaluation will be surveyed instead of 20 districts due to contamination by other education programmes or projects.

Annex B Supplementary information on EQUIP-T

B.1 Constraints acting on children’s ability to learn

Figure 23 Constraints on children’s capability to learn to their full potential

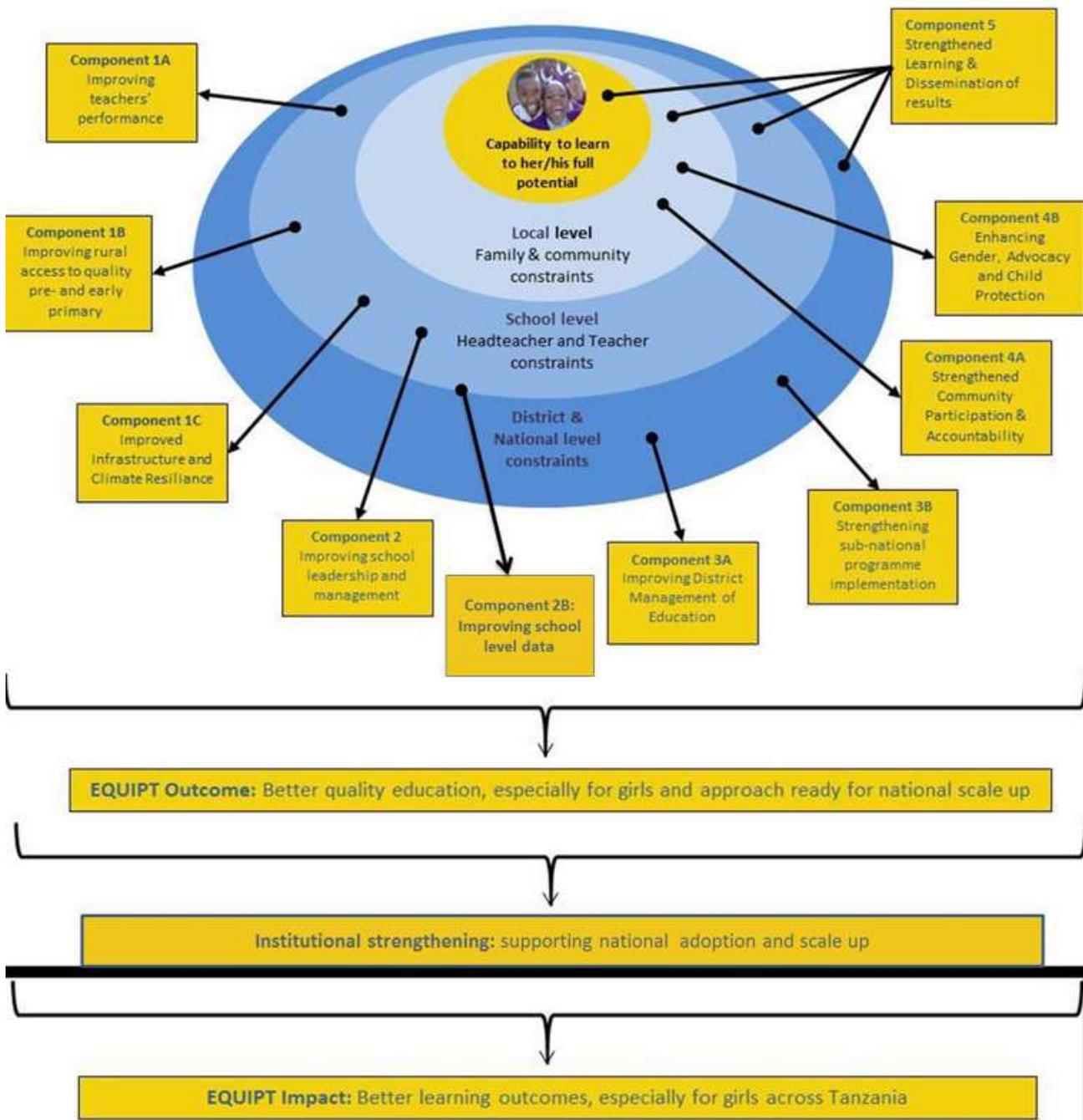


Source: Cambridge Education 2014, p6

B.2 Updated EQUIP-T theory of change

The updated EQUIP-T TOC set out below is taken directly from the EQUIP-T MA’s proposed programme extension document (EQUIP-T MA, 2017). This updated TOC was the starting point for the workshop with the EQUIP-T MA staff in January 2018, which aimed to elucidate more details on the underlying results-chains implicit in the summary TOC below. The result of this consultative process are the detailed results chains embedded in the endline evaluation matrix part I (Annex C)

Figure 24 Updated EQUIP-T theory of change (EQUIP-T MA)



B.2.1 Evolution of EQUIP-T’s theory of change

The failure of the Tanzania education system to deliver quality education was evident in the low levels of learning outcomes when the programme began the inception phase in 2013. EQUIP-Tanzania identified a number of barriers which prevented boys and girls from learning to their full potential. These focused on the provision of quality education as enrolment in Tanzania was already seen as a success story. The barriers were at community, school and governmental level (ward, district, regional and national). From these barriers EQUIP-Tanzania originally identified five critical medium-term, transformative changes that were required if the long term impact of better learning outcomes for all children is to be achieved. These are represented in the diagram above as components 1A, 2A, 3A, 4A and 5.

During the first 3 years of the programme, a deeper understanding these barriers has evolved. In particular, there is now an awareness that more than several million remote and marginalised children still lack access to education. As the programme has developed, a more nuanced understanding of the context, through EQUIP-Tanzania's adaptive management approach has enabled the programme to evolve to meet newly identified challenges. This has led to the revision of the theory of change and the addition of components 1B, 1C, 2B, 3B and 4B which are explained below. Also there is a greater understanding that to move from positive programme outcomes, to national level impact, a greater focus on institutional strengthening and national advocacy is required.

B.2.2 Assumptions and explanations behind the new elements in the theory of change

Output 1B –Improving rural access to quality pre- and early years primary: Evidence has emerged from UNICEF during the programme implementation period that 57% of 6 year olds and 67% of 5 year olds do not go to pre-primary. This proportion is higher in more remote areas and areas where Kiswahili is less widely spoken at home. Therefore the innovative School Readiness Programme has been developed to support more of these OOSC into pre-primary and support them to continue into primary school.

Assumption: if children are provided with quality pre-primary provision, where they learn Kiswahili they are more likely to enter Standard 1 and also more likely to learn effectively and progress when they get there.

Output 1C –improved infrastructure and climate resilience: poor infrastructure has always been acknowledged as a barrier, but two specific elements have contributed to it being included as a particular extension focus. Firstly, as explained above the number of rural out-of-school children without access to a nearby school calls for investment in satellite school infrastructure to reach these marginalised children. In addition, within existing schools, the spike in early years' enrolment during 2016, with average standard one class sizes increasing by 40% between baseline and end line highlights this issue as an impediment to education quality. There is an opportunity to support these schools and communities to become more climate resilient. Many of the rural, disadvantaged areas EQUIP-Tanzania supports are vulnerable to climate related shocks, such as floods, droughts, food shortages and storms. All of these factors have direct and indirect impacts on local children's education as well as their broader lives. The programme is proposing to support both greater awareness of local impact and mitigation strategies and to provide funding for adaptation and innovation through school grants to support schools to become more climate resilient in relation to water, energy and other issues.

Assumption: with better infrastructure class sizes can be reduced and the learning environment improved, leading to better opportunities for faster learning. With greater practical focus on climate change mitigation and resources to back this up pupils, schools and communities will be better placed to respond to challenges they face.

Output 2B –improving school level data: the lack of reliable and timely data from schools and the poor use of existing data for planning and performance management, has led to the programme developing a pilot School Information System. This system is designed to support better, more continuous data that is used for more evidence-based school performance management at various levels –this includes the ability for ward, district (including Quality Assurors), regional and national levels to be able to use the data to monitor quality, affect activities and even influence policy decision making.

Assumption: with better data, Head Teachers can target resources towards the areas that most need them and progress can be made more quickly and WECs can better assist them on this. In addition LGA, Quality Assurance, Regional and National officials can monitor where improvements are needed and also identify where progress is being made and share key lessons learnt.

Output 3B –strengthening sub-national programme implementation and ownership:as part of its adaptive management approach EQUIP-Tanzania decentralised a significant proportion of programme funds directly to LGAs, to increase ownership of quality improvement activities. This has been successful in enhancing ownership and also provides an opportunity to enhance sub-national PFM at the same time.

Assumption: continued decentralised funds will enable LGAs to play an ever greater role in the planning and management of these funds and the ownership of improvement and therefore be in a better position to understand and continue the most effective actions beyond the life of the programme.

Output 4B – enhancing gender equity: girls’ transition to secondary, disability and marginalised children: EQUIP-Tanzania’s approach to gender-equity so far has involved a number of mainstreamed approaches and a limited number of focused, gender-equity interventions. However, there is a clear need for a greater focus on some of the biggest barriers to girls maximising their potential. These include girls’ transition to secondary school and the levels of violence to which they are subjected.

Assumption: if some of these barriers are tackled pupils, particularly girls, disabled and marginalised children, will feel safer, come to school more often, stay in school longer, and therefore be able to concentrate on their learning development –ultimately leading to more pupils successfully transitioning to secondary schools.

B.3 EQUIP-T implementation progress between baseline and end-March 2018

EQUIP-T’s annual reports set out implementation progress. The last two annual reports (covering 2016 and 2017) provide summaries of progress under each component, but are less detailed than the earlier reports. Using these reports, together with information from interviews carried out with the EQUIP-T MA in January 2018, the tables below describe, for each component, the main activities implemented in the last two years (from midline to endline) in the seven regions. For completeness, implementation progress in the first two years of the programme (baseline to midline) is also included. Each table also contains a column called ‘ref’ which gives the original activity numbering set out in the EQUIP-T annual report 2015. This is useful because this it gives a picture of which activities have continued, are new, have been discontinued or were never implemented.

Table 21 Implementation of sub-component 1A teacher performance

Stage of impact evaluation		Baseline to Midline Aug 2014 to Mar 2016	Midline to Endline Apr 2016 to Mar 2018
Activities ¹	Ref	Description	Description
Early grade (EG) teacher INSET (district/ward level & school-based)	1.2	EG literacy modules 1-8; EG literacy modules 9-13 (partial); 3Rs curriculum INSET	EG literacy modules 9-13 (finish); EG numeracy modules 1-4; gender-responsive pedagogy (GRP); EG numeracy modules 5-9 ²
Communities of learning	new	X	Ward cluster reflection meetings for teachers related to INSET; school performance management meetings (SPMMs, aka 'weekly meetings')
Distribution of teaching and learning materials	1.2	Supplementary readers; big books/teacher read-aloud books; literacy toolkits	Numeracy toolkits
Teacher training videos	new	X	Production phase
Wider reform to teacher development		Various—see below	Institutional strengthening/partnerships with TIE, TSC, TTCs/Universities
Positive & safe learning environment campaign	new	X	Posters and activities aimed at teachers and parents (national campaign); video on positive learning environment in production phase
Teacher competency framework (TCF)/Teacher performance management system	1.1 /1.3	Produced TCF and concept note on teacher professional development strategy (TPDS). Intention to integrate simplified TCF into TPDS and operationalize in 2016	Contributions to National teacher CPD framework (at draft stage) ³

Sources: OPM 2016a, pp11-12; EQUIP-T MA (2016); EQUIP-T MA (2017); Interviews with EQUIP-T MA (Jan 2018). Notes: (1) Two of the original activities under this component, namely 1.4 Improving teacher morale, and 1.6 TTC scholarships for rural candidates, were not implemented. Instead, the intention (as noted in EQUIP-T MA, 2015) was to achieve the objective of 1.4 implicitly via support to the national teacher CPD framework; and for 1.6 to explore developing SRP Community Teaching Assistants into rural teachers under the SRP activities. (2) The EQUIP-T early grade numeracy training for teachers includes 13 modules overall, however, only the first nine out of the 13 were rolled out by mid-2018 due to delays in approval. (3) The latest two EQUIP-T annual reports (2016, p36 and 2017, p35), both emphasise the importance of a national framework for teacher CPD, to support the investments that have been made in INSET. The reasons for the delay in finalising and implementing this framework are not explained.

Table 22 Implementation of sub-components 2A and 2B school leadership and management

Stage of impact evaluation Timing		Baseline to Midline Aug 2014 to Mar 2016	Midline to Endline Apr 2016 to Mar 2018
Activities	Ref	Description	Description
School quality framework; Leadership competency framework	2.1	Both draft frameworks approved.	X ¹
SLM training & inclusive SDPs	2.3	HT/WEO training on 2 SLM modules; Partial roll out of HT/WEO training on SDPs (SLM 3)	Finishing SDP training, plus 2 review days on SDPs for HTs; ¹
School leadership communities of learning (HT COL)	2.4	X	HT/WEO training on SPMMs (aka weekly meetings); Training of trainers for district officials (DEOs/SQAs) on HT COL ² ;
Tablets for SIS	2.2	X	Tablets, sim cards and solar chargers, delivered to HTs and WEOs (some missing)
SIS training (school level)	2.2	Introduced in SLM 2 module	SIS launch/phase 1 training on SIS management for HTs and WEOs

Sources: OPM 2016a, pp11-12; EQUIP-T MA (2016); EQUIP-T MA (2017); Interviews with EQUIP-T MA (Jan 2018). Notes: (1) ADEM materials have absorbed/superseded the EQUIP-T developed frameworks under 2.1.; (2) The school leadership COL will be based on meetings convened by small groups of HTs (WEO not an active member but may monitor). The teacher COL will include the SPMMs, but may also include additional teacher meetings (no detail in existing EQUIP-MA annual reports on this, but highly unlikely to be implemented by time of EL survey).

Table 23 Implementation of sub-component 4A community participation and accountability

Stage of impact evaluation Timing		Baseline to Midline Aug 2014 to Mar 2016	Midline to Endline Apr 2016 to Mar 2018
Activities	Ref	Description	Description
Community based performance monitoring	4.3 4.10 4.12	School notice boards & support materials delivered (near complete roll out)	Development of community score card & piloting; District communications training (stories of change); District communication fund ³
Community engagement in education planning ¹	4.4 4.13	CSO facilitators trained Community Facilitators (CFs) to support community education needs assessments (CENA) & action plans	Finishing CENA & action plans in some communities
Capacity for effective operations of School Committee	4.5	SC training on roles, resp., school improvement, PTP set up; Separate SC training on PTP grants; TZS 450k grant (part of PTP grant#1)	SCs & PTPs training on roles of SC & PTP, policy, laws, resource management, school mission & vision
PTP formation & operations	4.7	PTPs formed, overseen by SCs. No direct training. TZS 100k from PTP grant#1 for PTP activities	PTP refresher training; PTP grant#2 TZS 550k for girls' education activities ²
School income generating activities (IGAs)	4.8	X	Training on business plan development for teachers and community business leaders; IGA grants of TZS 1.5m disbursed to 50% of schools per district, and implementation of IGA started

Sources: OPM 2016a, pp11-12; EQUIP-T MA (2016); EQUIP-T MA (2017); Interviews with EQUIP-T MA (Jan 2018). Notes: (1) There is an additional Activity 4.6 listed in the MA EQUIP-T report 2015, about linking CENA to SDPs in SDP training materials; it is given a status of 'achieved but ongoing'. (2) The EQUIP-T MA clarified that children with disabilities and other marginalised children could also benefit from the grant. (3) The procurement and delivery of noticeboards for schools which have not yet received them was delayed. The community score cards are called Community Based Performance and Monitoring (CBPM). These are distinct from the School Summary Report Cards that are being prepared by SQA teams following a Whole School Visit, and displayed on school notice boards.

Table 24 Implementation of sub-component 4B conducive learning environments for girls, children with disabilities and other marginalised children

Stage of impact evaluation Timing		Baseline to Midline Aug 2014 to Mar 2016	Midline to Endline Apr 2016 to Mar 2018
Activities	Ref	Description	Description
School clubs and child protection (CP)/anti-violence	4.9	Pilot in Mara to establish school clubs (JUU clubs) focusing on gender equity	WEO/HT/Tchs/PTP parents' training on JUU clubs; JUU clubs set up (Std5-7 girls/boys) with broader inclusion remit ¹ ;
	new (CP)	X	CP initiative at design stage; some schools have suggestion boxes for pupils to report issues; link with positive & safe learning campaign noted under sub-component 1A
Girls' education ^{2 3}			
PTP grant #2 for girls' education	new	X	See details under sub-component 4A above (PTP formation and operations)
Behaviour change communications	4.11	X	Shujaaz magazine ⁴ (monthly for 1 yr) for pupils, parents, communities; radio programme

Sources: OPM 2016a, pp11-12; EQUIP-T MA (2016); EQUIP-T MA (2017); Interviews with EQUIP-T MA (Jan 2018). Notes: (1) For example, disability, child protection, WASH. (2) Two other girls' education activities that are either under development or have not yet reached school level: (i) Gender-responsive toolkit; (ii) FGM initiatives in 3 regions. These have been excluded for brevity. (3) There is also a new girls' education activity that is beyond the scope of the IE: a secondary school readiness programme for girls which was piloted in 25% of wards in the EQUIP-T regions in Q4 2017. (4) Shujaaz magazine is a colourful comic-style publication that focuses on a young girl and the challenges she has to attain education.

B.4 EQUIP-T in-service training implementation details

Description of the EQUIP-T teacher in-service training delivered since baseline
<p>Objective</p> <p>To improve the performance of teachers, with a focus since baseline on strengthening early grade teaching of Kiswahili literacy (reading and writing) and numeracy and developing effective and gender-responsive pedagogy.</p>
<p>Delivery model</p> <p>A continuous professional development cycle that starts with residential training at the district level targeted at in-service training coordinators (INCOs) (each school appoints a senior teacher for this role), and sometimes includes head teachers, WEOs and teachers, delivered by a district in-service training team of teacher training college tutors. Following this, INCOs and sometimes teachers who attended the district-level training facilitate bi-monthly school-based in-service training sessions using group self-study and peer learning methods linked to classroom practice. Schools decide on the participants in school-based training but all teachers of Standards 1 and 2 are included at a minimum. Each study session takes about three hours and covers one module. Following this, the INCO and another teacher attend a ward cluster meeting each quarter with teachers from other schools in the ward. Teachers are also expected to attend school performance management meetings (SPMMs) each week with other teachers in the school to reflect on their classroom practice, and to get peer support and mentoring. There have been continued refinements to the model over time as the programme was learning what works best for improving the skills and knowledge of teachers and for ensuring the sustainability of the model.</p>
<p>Four sets of in-service training were provided for early grade teachers in 2015</p>

- *Set 1: Early grade Kiswahili literacy modules 1–4:* These cover general pedagogy, an introduction to gender-responsive pedagogy, and classroom management techniques. One day of ward-level training was delivered to INCOs and teachers of Standards 1 to 3, followed by school-based training.
- *Set 2: Early grade Kiswahili literacy modules 5–8:* These technical modules cover parts of the Kiswahili syllabus (reading and writing). They were delivered to early grade teachers as part of school-based training, followed by one day of district-level training for teachers of Standards 1 and 2, as a refresher.
- *Set 3: Early grade Kiswahili literacy modules 9–13:* These continue the series of technical modules covering parts of the Kiswahili syllabus (reading and writing). Three days of district-level training were delivered to INCOs and teachers of Standards 1 and 2, followed by school-based training. This was partially delivered in 2015 and completed in 2016.
- *Set 4: 3Rs curriculum training:* This covers the new Standards 1 and 2 national curriculum, including how to prepare schemes of work and lesson plans. Three days of district-level training was delivered to some teachers of Standards 1 and 2, followed by school-based training for the other teachers of Standards 1 and 2. In the rest of the country, 3Rs curriculum orientation training has been delivered by the LANES programme using a different model of one-off residential training.

Two sets of in-service training were provided for early grade teachers in 2016

- *Set 5: Early grade Kiswahili literacy modules 9–13:* These were partially delivered in 2015 and completed in 2016.
- *Set 6: Early grade numeracy modules 1–4:* These technical modules cover parts of the maths syllabus. Five days of district-level training were delivered to INCOs and specialist maths teachers (any maths teacher from Standards 1–6 selected by the school), followed by school-based training for Standards 1 and 2.

Two sets of in-service training were provided for early grade teachers in 2017

- *Set 7: Gender responsive pedagogy:* This covered gender-responsive pedagogy. Three days of district-level training were delivered to INCOs and gender-focal teachers (if schools had teachers appointed to that post), followed by school-based training for teachers of all Standards.
- *Set 8: Early grade numeracy modules 5–9:* These continue the series of technical modules covering parts of the maths syllabus. Five days of district-level training were delivered to INCOs and specialist maths teachers, followed by school-based training for Standards 1 and 2.

Sources: EQUIP-T MA (2015); EQUIP-T MA (2016); EQUIP-T MA (2017); Interviews with EQUIP-T MA (Jan 2018).

B.5 EQUIP-T SLM implementation details

EQUIP-T component 2: strengthening SLM implementation

Implementation by midline (Aug 2014 to Mar 2016)

- School quality and Leadership competency frameworks developed and approved.
- HT / AHT / WEO training on **(the expected number of training days shown here are indicative and may vary by district)**:
 - SLM module 1 (3 days): education quality standards and responsibilities and roles of head teachers.
 - SLM module 2 (3 days): how to lead EQUIP-T school-level initiatives. Topics covered PTP grant management, SRP set-up, SIS introduction, and school development activities, including communication.
 - SLM module 3 (3 days): school development planning (**partial rollout**).

Implementation by endline (Apr 2016 to Mar 2018)

- HT / WEO training on **(the expected numbers of training days shown here are indicative and may vary by district)**:
 - SLM module 3 (3 days): school development planning (**completion of rollout**).
 - Review of school development planning (1+1 days, only for HTs).
 - School performance management (3 days): SPMMs, ward education meetings and developing school mission and vision.
- SIS phase 1 on SIS management (5 days).
- Provision of SIS tablets to all head teachers.
- Training of trainers (DEOs / SQAs) on communities of learning.

- Communities of learning (COL) training for districts.

Sources: EQUIP-T MA (2017), OPM (2016a), OPM (2018).

B.6 EQUIP-T logframe indicators

The indicators in Box 17 are taken from the latest logframe made available to the evaluation team (dated November 2017). Generally, there has been a lot of change in logframe indicators over the course of the programme.

Box 17 List of EQUIP-T logframe indicators related to the components covered by the first part of the endline evaluation

EQUIP-T logframe indicators related to teacher performance (Component 1A)
Average early grade teacher performance in pedagogy ¹
Percentage of teachers using gender-responsive pedagogy in their classroom teaching
Percentage of schools where teachers are engaged in active in-service training groups or COL
EQUIP-T logframe indicators related to SLM (Components 2A and 2B)
Percentage of head teachers achieving a target number of the following performance management indicators ² :
1. School plan completed with gender responsive strategies.
2. School plan implementation started and progress tracked.
3. Weekly management meetings with teachers.
4. In-service training sessions at least once a month (minimum six sessions per term).
5. School cash book in place and updated each quarter.
6. School committees elected and meet at least quarterly.
7. School club in place and active.
8. PTP in place and active.
9. SIS data updated (there is both male and female data).
10. Follow up on teacher attendance from SIS data.
Percentage of schools implementing five or more activities from the SDP
Number of schools with pupil data uploaded onto SIS.
Number of wards with active COLs
EQUIP-T logframe indicators related to community participation and demand for accountability (Component 4A)
Percentage of schools reporting changes made as a result of meetings based on scorecards
Percentage of schools with parental engagement in school improvement through active: a) PTPs; b) SCs
EQUIP-T logframe indicators related to conducive learning environment for marginalised children (Component 4B)
Percentage of schools with PTPs participating in activities to improve inclusion or child protection in their school ³
Proportion of teachers using gender-responsive pedagogy in their classroom teaching
Percentage of schools with a) at least one school student welfare club established and b) club engaging in at least one activity specifically related to girls' welfare or gender responsive measures
Source: EQUIP-T MA, 2017

Notes: (1) This indicator measures teacher's overall pedagogy score which is defined as a percentage. It is based on observing teachers' demonstration of 31 different competencies during classroom observation, each of which has a different pre-determined weight in the overall score, and then dividing the total score achieved for each teacher by the total possible sum of scores. (2) The target number of strategies changes each year: six (2017), seven (2018) and eight (2019). (3) This indicator is measured by three criteria: PTP in place, PTP active, and PTP "involved in reporting cases of breach of child protection or been involved in providing information to others on addressing child protection (such as local campaign, or information to students about their rights) or has been involved with activities for inclusion"

Annex C Endline evaluation matrix, part I

Table 25: Endline evaluation matrix part I (partial, related to quantitative evidence)

Prog. Comp.	Results chain level	Descriptive link in theory of change	Endline evaluation questions (partial, to be answered using quantitative evidence) ¹
All	impact	Better learning outcomes, especially for girls Contributions from all component result chains reinforced by interlinkages. Strong governance model (community monitoring to school leadership to ward co-ordination to district management to regional strategic leadership), acts as a foundation for impact and sustainability.	Did standard 3 pupil learning in Kiswahili and maths improve? Why? To what extent can this be attributed to EQUIP-T? Did learning gaps narrow for marginalised groups (girls, non-Kiswahili speakers)? Why?
All	outcome	Better quality education, especially for girls (positive, inclusive and girl-friendly learning environment in schools); improved retention rates for girls and boys ²	No specific additional questions related to 'better quality education, especially for girls' beyond those captured under the components below; all components contribute to this outcome, and the evidence will be synthesised to get a holistic picture.
All	intermediate outcome to outcomes and impact (assumption)	DEMAND-SIDE SUPPORTING CONDITIONS RIGHT: regular pupil attendance; school-ready children/attendance at pre-school; adequate support at home for school work; positive parental attitudes to education for girls/CWD/OVC; other social-economic conditions do not pose overwhelming barriers; appropriate class sizes	Did assumptions on pupil attendance, pre-school attendance, support at home, hold? Have there been changes over time? Has pupils' household poverty status changed over time?
Teacher performance			
C1A: Tch	input	INSET for early grade (EG) teachers; teaching and learning materials (TLM) provided; positive/safe learning campaign materials provided; info. on structures for teachers' communities of learning (COL) provided	Did EG teachers receive district/ward INSET as intended? Did school-based INSET happen as intended? Did school receive materials as intended?
C1A: Tch	output	EG teachers' capacity, confidence increased; TLM & positive learning campaign materials available in classrooms; COL structures operating	Did EG teachers' capacity (curriculum knowledge, general/inclusive pedagogical skills) increase? Are teachers more confident? Are materials in classrooms? Are teacher COL structures operating?
C1A: Tch	input to output (assumption)	Teachers readiness to learn; relevance/accessibility of INSET & TLM materials; teacher attendance at INSET; quality of INSET maintained from district/ward level to school level; gender-based attitude/ customs are open to change	Did assumptions on INSET attendance hold? What challenges, if any, do teachers face in attending and learning from INSET?

Prog. Comp.	Results chain level	Descriptive link in theory of change	Endline evaluation questions (partial, to be answered using quantitative evidence) ¹
C1A: Tchs	intermediate outcome	Classroom teaching improved; materials being used effectively; teachers using positive behaviour management; instructional hours appropriate	Has classroom teaching improved (more active, inclusive, gender-responsive & use of TLM)? Are teachers using more positive behaviour management? Have instructional hours increased?
C1A: Tchs	output to intermediate outcome (assumption)	SUPPLY-SIDE SUPPORTING CONDITIONS RIGHT including high teacher school & classroom attendance; punctuality; low teacher turnover; limited movement of EG teachers to upper stds; high morale/motivation and job satisfaction	Did assumptions on attendance, punctuality, turnover, morale hold? If not, why? Have there been changes over time? Are supporting conditions for teachers (e.g. overall workload, outstanding payments due, housing etc.) improving?
School leadership and management			
C2A/B: SLM	input	HT/WEO training on SLM, school development planning (SDP), school performance management meetings (SPMMs), school information system (SIS); tablets delivered; info. on structures for HTs' COL provided; WEOs with transport and grants (from C3A)	Did HTs receive training as intended? Did schools receive tablets?
C2A/B: SLM	output	HT capacity (knowledge/skills) increased; annual SDPs that meet quality criteria available; SPMMs happen weekly & teachers share problems/solutions; HTs meet as COL & share problems/solutions; SIS is functional; produces useful reports to support SLM; effective monitoring/support from WEOs	Has HT SLM capacity increased in core areas (planning, tch & resource m/ment, community engagement, communication/reporting)? Are 'quality' SDPs available for the current year (evidence-based, costed, incl. teaching and learning & inclusive education activities)? Are SPMMs happening? How frequently? What is discussed? Are HTs meeting as COLs? Is the SIS up-to-date with pupil/teacher records? Is daily attendance being recorded? Can HTs produce useful analysis/reports from the SIS? Have WEO visits increased? Are there any changes in what WEOs are doing during visits?
C2A/B: SLM	input to output (assumption)	HT/WEC readiness to learn/ability to use technology; relevance/accessibility of INSET materials; SC/communities have time to be actively engaged in SDP; HT/WEO attendance at INSET; HT/teacher attendance at SPMMs; HT attendance at COL meetings; SIS software is well designed for purpose; low turnover of WEOs	Did assumptions on attendance hold? If not, why?
C2A/B: SLM	intermediate outcome	HT leads school more effectively by applying new skills and knowledge (positive changes in attitude and confidence); SDPs are implemented; HT takes action based on SIS; greater school transparency	Has SLM improved? In which areas? What actions has the HT taken to improve the school? Have some activities from SDPs been implemented? Does the HT use the SIS for SLM? Does the HT make public key information on pupil performance, teacher/pupil attendance & resource use?

Prog. Comp.	Results chain level	Descriptive link in theory of change	Endline evaluation questions (partial, to be answered using quantitative evidence) ¹
C2A/B: SLM	output to intermediate outcome (assumption)	High level of HT school attendance and punctuality; low turnover of HTs; HT morale/motivation and job satisfaction is high; monthly capitation grants received in full; little shortfall in planned resources receipts	Did assumptions on attendance, punctuality, turnover, morale/motivation hold? If not, why? Have there been changes over time? Are supporting conditions for HTs (e.g. overall workload, outstanding payments due, housing etc.) improving? Are capitation grants received monthly in the expected amounts? What other sources of funding does the school have?
Community participation and demand for accountability			
C4A Com	input	School committee(SC)/Parent-teacher partnership (PTP) training; PTP grant#1; business plan training; IGA grant; community needs assessment (CENA) conducted; notice boards distributed	Was training received as intended? Did schools receive grants? Did schools receive noticeboards?
C4A Com	output	SC capacity (knowledge/skills) improved; SCs active; PTPs set up & active; PTP grant spent; capacity for school-community IGA built; IGA grant spent; CENA action plan in place & linked to SDP; school noticeboards publically accessible with relevant information	Is the role of the SC and PTP clear and distinct? Do SCs meet regularly? What do SCs discuss? Have PTPs been set up, and are they active? Have the grants received been fully spent, and on what? Are HTs aware of CENA action plans? Do schools have publically accessible noticeboards? What is displayed on them?
C4A Com	input to output (assumption)	Training materials/approach are accessible/relevant; attendance at training/CENA process is high; SC/PTP members have time & motivation to change/engage; PTPs able to self-organise; no conflict between SC & PTPs; HTs open to increased transparency/scrutiny;	See footnote (1)
C4A Com	intermediate outcomes	Better engagement between schools & all parents/wider communities (mutual understanding, actions & support for school improvement); improved communication between schools & all parents/wider communities; parents/ communities hold schools to account	Have PTPs taken action to improve education in last school year? Are all parents aware of PTPs? Has an IGA been established, using the grant? Is the HT aware of any activities resulting from the CENA action plan? Do parents read the school noticeboard? Do parents receive regular information on the progress of their children? Do parents receive information in other ways?
C4A Com	output to intermediate outcome (assumption)	PTP & community members have time, resources & motivation to support schools; free education policy does not undermine community motivation; parents visit schools regularly; no misuse of funds; no political interference	See footnote (1)

Prog. Comp.	Results chain level	Descriptive link in theory of change	Endline evaluation questions (partial, to be answered using quantitative evidence) ¹
C4B Com	input	Tchs/PTP members training on JUU clubs; PTP grant#2 for girls' education; Shujaaz magazine & radio campaign	Did JUU club training take place? Did schools receive the grant? Did school receive Shujaaz magazines?
C4B Com	output	JUU clubs set up and active; PTP grant#2 spent; Shujaaz magazine distributed to the target group	Do schools have JUU clubs? Do they meet regularly? Has the grant been spent?
C4B Com	input to output assumptions	Training materials/approach are accessible/relevant; attendance at training is high; little resistance in the school community to increasing openness around taboo subjects; PTP members have time & motivation to plan grant activities for girls' education	See footnote (1)
C4B Com	intermediate outcome	Better support for girls, & other marginalised groups incl. CWD, OVC, to be safe, attend school regularly, and learn; better community engagement & support for education for girls & other marginalised children	What actions have the JUU club taken in the last school year? What was the PTP grant#2 spent on? What are the main issues discussed between the school and all the parents when they meet?
C4B Com	output to intermediate outcome assumptions	little resistance in the school community to increasing openness around taboo subjects and taking related action; no misuse of funds; no political interference	

Source: OPM 2018a. Notes: (1) The blank cells in the column under Component 4 mean that no quantitative evidence will be collected related to the assumptions described. These assumptions have been included for completeness, but require qualitative methods to capture (as was done at the midline). (2) Retention rates are not captured in the quantitative IE survey. The IE team will explore EMIS data (publically available on the government's opendata website) to see if any trend data in primary school retention rates for the 5 original EQUIP-T regions (and the 17 districts within these regions that the EQUIP-T IE survey estimates represent) is available, and could be presented as additional context for the evaluation findings.

Annex D Education policy context and other programmes

Education policy changes since baseline

The Ministry of Education, Science and Technology (MOEST)¹⁷⁴ leads in setting education policy in Tanzania. Since 1995, the system has been organised under an overarching policy document (the Education and Training Policy (ETP)) which sets the aspirations and structure for education. Under the ETP I, the Ministry developed two consecutive ten-year ESDPs which are now at the end of their implementation period. Under these ESDPs, there have been five-year sub-sector plans, such as the Primary Education Development Programme (PEDP I, II and III) and the same for secondary (SEDP I and II). Under these there are donor-supported programmes.

A new Education and Training Policy (ETP II, 2014) was launched in February 2015, and this together with other national planning documents (Tanzania Development Vision 2025, and the National 5-year Development Plan) provided the foundation for the latest ESDP 2016/17 to 2020/21. This 5-year plan for the education sector, has been endorsed by the Education Sector Development Committee (ESDC), and is awaiting finalisation. The key policy initiatives of most relevance to the objectives of the EQUIP-T programme are:

- Universalisation of basic education:** provision of 11 years of free and compulsory basic education to all children—this spans pre-primary (1 year), primary, and lower secondary (forms 1 to 4).
- Equitable access for all, including the most disadvantaged children:** a focus on tackling various forms of disadvantage related to, for example, gender, location, having a disability, being an orphan or being vulnerable for other socio-economic reasons.
- **Improved quality of basic education,** via three main strategies: (i) investment in teachers, (ii) tracking curriculum competencies attained by students; and (iii) a revamped school quality assurance (SQA) system.

The **policy on free school education** was one of the first actions to be **implemented** by the new government in **December 2015**. The MoEST released a circular that states that parents and guardians will not have to pay for education of their children from standards one to form four.¹⁷⁵ Part of the circular reads: '*Provision of free education means pupils or students will not pay any fee or other contributions that were being provided by parents or guardians before the release of new circular*'. The directive does not explicitly mention pre-primary, but the new ESDP is clear that pre-primary is also free for pupils. This change in cost burden away from parents would be expected to increase demand for basic education, and thus to raise enrolment particularly at pre-school and standard 1 in the short-term.

Capitation grants: To replace parental contributions (voluntary contributions at pre-primary/primary, and fees/contributions at lower secondary), public schools now typically rely entirely on government funding for running costs. Prior to the school year 2016, schools received capitation grants from the government via district authorities. Payments were extremely erratic, and typically less than the prescribed norm per pupil. The new government changed the funding modality, so that from January 2016, capitation grants¹⁷⁶ flow directly from the central Ministry of Finance to school bank accounts on a monthly basis. This change has resulted in schools receiving monthly capitation grants (and typically

¹⁷⁴ Prior to the new government which took office in November 2015, the Ministry had responsibility for vocational training and was known as Ministry of Education and Vocational training (MoEVT).

¹⁷⁵ Government Circular Number 5 of 2015

¹⁷⁶ The term 'capitation grant' is an umbrella term for various elimu bure (free education) payments, including capitation grants for primary and secondary schools, subsidies for secondary school fees, and subsidies for special needs schools.

the same amount each month).¹⁷⁷ The Education Program for Results (EPforR) discussed in the next section, incentivises the regular payment of capitation grants.

Competency-based early grade curriculum: There has been a significant change in the early-grade primary school curriculum, both what is being taught and the pedagogical approach. In 2015, the government introduced a new competency-based curriculum for standards 1 and 2. Subjects were drastically reduced to focus on reading, writing and arithmetic skills (3Rs).¹⁷⁸ The new curriculum, syllabi and teachers' guides promote a phonics approach to teaching children to read, which was new to most teachers. Since 2016, the government has introduced a revised curriculum for standards 3 and 4, which follows on from the standards 1 and 2 curriculum in approach, but has more advanced content. The Literacy and Numeracy Education Support programme (LANES) has been a key driver of curriculum reform, and this programme, together with others, including EQUIP-T, is providing materials and in-service training (INSET) to support teachers. More details are in the sections below.

School quality assurance (SQA): The MoEST's SQA division leads this initiative. The reformed process of SQA (previously traditional school inspection) is nearing completion, and is intended to be outcome based, more collaborative, and comprise both internal school-level SQA and external SQA (involving SQA officers and others). The quality assurance operational manual, and handbook, are currently being finalised (MoEST 2017c and 2017d). Whole-School Visits (WSVs) are expected to start well before June 2018, following orientation of SQA officers.¹⁷⁹ The EPforR (discussed below) incentivises WSVs and the public display of school summary report cards.

Human resource management: Two government initiatives implemented since 2016 are relevant to key education professionals. First, the government introduced higher qualification requirements for head teachers and for Ward Education Officers (previously Ward Education Co-ordinators), as part of a professionalization initiative. This has meant that a sizable number of head teachers and WEOs have had to leave their positions.¹⁸⁰ Second, the government clamped down on civil servants, including teachers, with fake qualifications which resulted in a large number of dismissals.

¹⁷⁷ The flow of capitation grants to primary and schools was assessed as part of the EPforR annual verification exercise 2016/17.

¹⁷⁸ 24 out of 30 periods per week are allocated to 3Rs, leaving 6 periods for supportive skills (health and environmental education; games and sports; fine and performing arts; and religious studies)

¹⁷⁹ The training of trainers took place in January 2018, but by late February 2018 the orientation of SQA officers had not yet started.

¹⁸⁰ There may also have been a similar change in qualification requirements for HTs.

Annex E Teaching practices descriptors

Table 26 Teaching practices and descriptors

	The teacher	Teaching practice descriptor
Lesson introduction		
1	States the objectives of the lesson, and introduces the topic in a clear way	Learning objectives are clearly stated at the beginning of the lesson. Teacher explanation is accurately and clearly presented with good signposting and makes strong connections to pupil experience.
2	States what new skills or knowledge pupils will have by the end of the lesson	Teacher specifically states what new skills or knowledge the pupils should have acquired by the end of the lesson. For example, solving particular type of problems in maths or a specific writing skill in Kiswahili.
3	Checks for prior knowledge of the topic among the pupils	Teacher asks pupils about previous work covered in the topic and questions them about their understanding.
Lesson middle stage		
4	Asks pupils to demonstrate in front of class	Teacher calls on pupils to answer questions, explain ideas and report back on activities in front of the class.
5	Asks open-ended questions	Teacher asks questions which have more than one answer. Teacher asks questions which encourage speculation and require more than a 'yes' or 'no' answer or the recall of information.
6	Probes or comments on pupils' answers	Teacher asks the pupils for further explanation of his/her answer (probe). Teacher uses pupils answer to give an example, or expands, or provides additional information (comments).
7	Encourages pupils to ask questions	Teacher encourages pupils to ask questions to the teacher or to other members of the class.
8	Provides written or verbal feedback to pupils on their individual work	Teacher provides spoken comments to pupils individually on their work. Teacher provides written feedback such as marking of work, including formative feedback if the pupil has made mistakes or does not understand well.
9	Uses paired or group work	Pupils carry out activities in pairs or in groups.
10	Makes effective use of the chalk/white board	Teacher's writing and diagrams are clearly laid out.
11	Uses different instructional materials	Teacher makes use of a variety of instructional aids (not the blackboard) such as maps, posters, tables, charts, real-life items.
12	Relates well to pupils and uses praise	Teacher conveys enthusiasm through voice and body language. Teacher has a good rapport with pupils. Teacher uses encouragement and praise to give positive feedback. Teacher calls on pupils by name to make a contribution to the lesson.
13	Switches between Kiswahili and a vernacular language	Teacher code-switches between Kiswahili and a vernacular language during the teaching and learning process.
Lesson end stage		
14	Checks if pupils have acquired the new skills or knowledge stated in the introduction	Teacher asks questions or uses another approach to find out if pupils have acquired the new skills or knowledge set out in the introduction.
15	Uses a plenary (whole class session) to summarise and extend learning	Teacher draws the whole class together at the end of the lesson to summarise what has been covered in the lesson; consolidate and extend learning by directing pupils to the next stage of learning.
Source: OPM (2014b).		

About the project

The independent Impact Evaluation of the Education Quality Improvement Programme in Tanzania (EQUIP-T) is a 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 after the programme ends; and iv) communicate evidence generated by the impact evaluation to policy-makers and key education stakeholders.

EQUIP-T is a six-year Government of Tanzania programme, funded by UK DFID, which seeks to improve the quality of primary education in nine regions of Tanzania, and thus to improve learning outcomes, particularly for girls. It focuses on strengthening performance of teachers, school leadership and management, systems which support district management of education, and community participation in education.



Oxford Policy Management

