

# EQUIP-Tanzania Impact Evaluation

## Final Endline Report

Nicola Ruddle and Georgina Rawle

April 2020

## 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 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 and providing comments on drafts of previous evaluation outputs.
- 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, midline, and endline evaluations: those who provided technical inputs to the design, and the field teams who collected the data, as well as the analysts and reviewers. Full acknowledgements can be found in the baseline, midline, and endline evaluation reports (Oxford Policy Management (OPM), 2015a, 2017a, 2019a, 2020a, and 2020b).
- Stephen Jones (OPM), who advised on the design of this endline report.
- The other authors of the endline technical studies: Michele Binci, Gunilla Pettersson Gelandar, Jana Harb, Paul Jasper, Safa Khan, Deo Medardi, Alessio Romarri, Michelle Rorich, Mehjabeen Jagmag, Kelly Casey, Peter Sutoris, and Gabi Elte; in particular, Kelly Casey, Jana Harb, Mehjabeen Jagmag, and Deo Medardi, who provided inputs to the initial development of the conclusions, recommendations, and lessons of this final report.
- Last but not least, all the studies' respondents: head teachers, teachers, pupils and their parents, national, and regional and district level government officials, who generously gave their time and shared information and views.

This report was reviewed by Professor Herme Mosha (University of Dar es Salaam), Paud Murphy (independent consultant, and former World Bank Lead Education Specialist), and Dr Caine Rolleston (University College London, Institute of Education).

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

## Executive summary

### Overview

**This is the final endline report of the impact evaluation (IE) of the Education Quality Improvement Programme in Tanzania (EQUIP-T), conducted by Oxford Policy Management (OPM). 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 planned to reach 2.3 million primary pupils by 2018, about one-quarter of total primary enrolment in Tanzania, in five (extended to seven in 2015) of the most educationally disadvantaged regions in Tanzania.<sup>1</sup> A contract extension in 2017 expanded the programme to nine regions, introduced new sub-components, extended the end date to January 2020, and increased the budget from approximately £50 million to £90 million. The programme has a managing agent (MA), which works with the government to deliver the programme, and is responsible for £80 million of the total EQUIP-T budget. The activities overseen by the MA are the focus of the IE.

The programme consists of five broad components: (1) improved access to high-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.

### Objectives and scope

The overall objectives of the EQUIP-T IE 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 IE to policymakers and key education stakeholders.

This final summary report draws on all the earlier evaluation rounds to present a final summary of the evaluation's findings and conclusions.

This report has two primary audiences: the Government of Tanzania, which is considering how to adopt and adapt activities from EQUIP-T and scale these up nationally; and DFID, for accountability and learning, particularly as it considers future education programming in Tanzania. More widely, the report is intended to be of use for education sector stakeholders in Tanzania and beyond in informing education programming. The Executive Summary aims to provide a succinct summary of the objectives, methodology, conclusions, recommendations, and lessons learned from the report.

### Methodology and evidence base

The endline forms part of a mixed-method evaluation that began with a baseline in 2014. This was followed by a midline in 2016, and then by two phases of endline studies: a quantitative endline study carried out in 2018; and then a qualitative study and a separate cost study, both conducted in 2019.

<sup>1</sup> There are 26 regions in mainland Tanzania.

The quantitative estimation of programme impact on pupil learning uses a quasi-experimental approach that combines propensity score matching with difference-in-differences. Quantitative survey data and qualitative studies, together with other secondary sources, are combined in order to clarify key channels of programme influence, or reasons for lack of change, addressing research questions structured around the programme's theory of change (TOC). The cost study provides greater breadth to the evaluation by responding to questions on the efficiency and sustainability of the programme.

The final endline report answers a series of evaluation questions to cover the entirety of the endline evaluation. The evaluation questions were drawn from the OECD Development Assistance Committee (DAC) evaluation criteria: relevance; impact; effectiveness; efficiency; and sustainability.

## Findings and conclusions

The key findings and conclusions are structured around the five DAC criteria: impact; effectiveness; relevance; efficiency; and sustainability.

### Impact

**EQUIP-T had a substantial positive impact on learning outcomes, contributing to the programme's overall objective.** In both Kiswahili and maths, the programme helped children in Standard 3 move up from low-performing bands and come closer towards meeting Standard 2-level competency. The impact was particularly notable for children at the bottom of the performance distribution, helping many more children move up from this very low base than would have been the case in the absence of EQUIP-T. The improvement was larger in Kiswahili than in maths, likely due to the longer focus on Kiswahili. Whilst the logframe targets were not quite achieved, the improved outcomes of the lowest performing pupils are significant and not captured in the logframe.

**This success is all the more notable given some of the contextual factors and changes taking place.** In particular, there were substantial levels of teacher and head teacher turnover, and growing class sizes; each of these were prompted by new government policies during the period of EQUIP-T and put downward pressure on teaching and learning. The increase in enrolment actually means that far more pupils are meeting curriculum standards in absolute terms. Furthermore, the change in curriculum in 2015 was complementary to EQUIP-T's focus on numeracy and literacy, and so supported EQUIP-T's aims. Meanwhile, the programme was implemented in regions with a difficult starting environment: a large share of households below the poverty line, low school quality, and large class sizes.

**Despite the large improvement in learning outcomes, it is critical to note that pupil learning outcomes are still very poor, and fall well behind curriculum expectations.** Furthermore, certain groups are particularly disadvantaged, specifically those who do not speak Kiswahili as their home language, and those from poorer households.

### Effectiveness

**The evaluation concludes that the main contributor to this impact was the teacher in-service training,** given it had the closest direct link to what children learn in the classroom and was implemented largely as planned. The other components were supporting factors to varying degrees, working through aspects of better district education management and school improvement activities initiated by grants, but none of the intermediate outcomes have seen unequivocal improvements.

## Teacher performance

**The teacher in-service training was clearly needed and has been effective.** Although improvement over a range of measures of effective teaching practice has been mixed, the in-service training has helped teachers improve their classroom management, given them a range of tools for use in the classroom, led to more use of teaching aids, and built teacher confidence and motivation. Based on the measures of positive teaching practices seen in the quantitative lesson observations, use of some has improved and yet others have significantly worsened. Teachers are using different instructional materials more frequently in their lessons. However, despite the distribution of teaching and learning materials (TLMs) to schools, EQUIP-T's materials are typically not found in classrooms and are not being used, so have not led to effective change.

The inputs and outputs for the in-service training were largely delivered as intended, with almost all teachers attending some in-service training. However, the implementation of the school-level training model has faced some challenges which could have been foreseen at design stage, such as complaints about the lack of allowances and difficulties fitting the sessions into the school day, yet despite this the sessions are taking place (albeit with wide variation). For the school-based model to be sustainable, there will need to be commitment and support from all levels of government, and some kind of incentive for teachers to take part, such as building in-service training into the career progression model.

As mentioned, staff turnover – from teachers through to Local Government Authority (LGA) staff – is very high and a major issue for the effectiveness of interventions which rely on capacity building and improved relationships. Learning and skills may be lost and the community of learning (COL) ethos may wane as original groups of colleagues disperse. Even at higher levels of government, the turnover in central ministries and regions affects the level of buy-in and thus leadership and the commitment to sustaining the EQUIP-T activities. A further underlying issue is the severe level of teacher absenteeism from classrooms and its effect on time on task (the amount of time pupils spend attending to school-related tasks), which is a major factor for children's learning. Whilst time on task was originally in EQUIP-T's logframe, there was never an explicit intervention to address this, and although improvements were seen, it continued to be a problem up to the end of the programme. High teacher turnover and limited time on task – both risks which could have been foreseen in the design phase and the level of which depend on wider systemic change – affect the delivery of the intended benefits of the in-service training, and also the sustainability of the training going forward.

## SLM

**At the time of this evaluation's data collection, the school information system (SIS) was not being used and thus would need to be assessed and improved in order to make it an effective tool for the use of data in school management** – it is possible some changes have already taken place since 2018. The reasons for lack of use were largely design issues – difficulty and time taken to enter data, and the tablets not working properly. However, poor internet connectivity is also a systemic issue which affects the effective use of the SIS. The other major challenge is that head teachers need sufficient training to use the system, and with the level of turnover, over half of head teachers had not received the training.

Changes in measures of school leadership were mixed, with schools holding more staff meetings and taking actions to address pupil attendance, but whilst head teachers are monitoring lesson plans, they are observing fewer lessons. Whilst only half of schools have a school development plan (SDP), most of those that do have started implementing it. The relatively low outcomes from this component are likely due to the high turnover of head teachers (intensified by the new qualification policy in 2017) and head teacher absenteeism, but they also raise the question of the effectiveness of the training content.

## District management

**In terms of district management, the programme has had little effect in the district office on education planning capacity and processes.** This is not a surprise given the programme's shift in focus towards supporting implementation through LGAs rather than capacity building, and in this regard the programme was able to implement a substantial volume of activities through government systems. There was more positive change in building the awareness and understanding of LGA officers of education issues, and in introducing district education meetings (DEMs) as a method for collaborative education management and raising capacity. DEMs are valued by LGA officers and Ward Education Officers (WEOs) and are a low-cost activity which could be continued, if encouraged.

**In LGAs identified as successful, EQUIP-T has empowered and built the capacity of WEOs** through various training opportunities, including DEMs. It has also facilitated WEOs to visit schools more frequently due to the provision of motorbikes and grants. WEOs' hard work and commitment has also been supported by the government's focus on work ethic, the provision of responsibility allowances, and the higher minimum qualification policy, all of which have elevated WEOs' position and morale. However, a wider sample of head teachers in 2018 were less positive about the support they receive from WEOs, which may relate to the high WEO turnover experienced in 2017.

## Community participation

**The community component has contributed to pockets of improvement in the relationship, support, and communication between schools and communities. However, on an aggregate level, there are still challenges and the general situation and levels of accountability have not improved substantially.** The assumptions required for the interventions to lead to greater accountability were not well defined in the programme design, and barriers include parents' lack of self-belief and the view that schools would not take action anyway. Ultimately, some of the underlying challenges – teachers' negative views of parents, who are seen as lazy and uninterested in their children's education; and parents feeling unable to approach teachers or hold the school to account – are substantial and will take a long time to change.

**Parent-teacher partnerships (PTPs) have not been successful at bringing the wider parent body closer to schools – they are not well known by parents who are not PTP members.** The space for engagement between schools and communities is led by school committees (SCs), which are legally established, but were not ready to share that space, and PTPs have not been empowered to play their role in supporting decision making and school improvement. However, at a lower level of ambition, parent members of PTPs in around two thirds of schools have been more actively engaged in school matters, bringing the core parent members of the PTP closer to the school and supporting student achievement. The PTP grants and income-generating activity (IGA) grant were instrumental in mobilising parents and community members, including engaging the PTP itself, and it is unlikely the PTPs will survive without these grants as catalysts. In successful cases, parents and the community did mobilise to support these projects with contributions, labour, and a sense of ownership.

Schools and communities are confused about the difference between PTPs and SCs, and in fact SCs became more active when they received training. Repeating the SC training will be important for their activity to be sustained. More generally, the issue of allowances or incentives for attending some school activities is likely to become a sticking point in any attempts to raise community involvement. Finally, the school noticeboards have not been effective at improving communication between schools and parents.

## Design and implementation

**Some of the challenges that have impeded the programme's effectiveness result from the underlying assumptions in the design; these should have been foreseen.** For example, the barriers to holding meetings (with teachers or community members) without incentives and fitting it into other schedules; or the loss of capacity due to high turnover. The challenge of changing cultural norms and behaviour could have been better recognised, particularly in the design of the community component. There are other examples across the components. Although this is not well documented, it is possible that these challenges were foreseen and considered tolerable. At the same time, the constraints on learning initially identified by the MA are substantial and EQUIP-T's interventions might be expected to make only incremental improvements given the complexities of the system and the length of time needed to effect change.

**However, other reasons for low effectiveness were caused by implementation failings.** Over time, the programme introduced many new activities, sometimes at the expense of others. For example, the shift to decentralised implementation led to less focus on public financial management (PFM) capacity building, and the introduction of the infrastructure sub-component required new expertise in the MA and monitoring systems in LGAs. It is possible that some of the later delays, such as the delayed sub-component 4B activities on inclusion and gender, which were part of the original design, were a result of trying to do too many things. Arguably, the MA spread itself too thin, without an evidence-based analysis of trade-offs.

## Relevance

**The design was likely to achieve results, but with some notable risks.** The original programme design was largely likely to achieve its results based on the strength of evidence that there were links between interventions, outcomes, and intended impact. However, there were a number of key risks in the assumptions which appeared likely to limit effectiveness: the necessary motivation to attend training and to change behaviours; high levels of education staff turnover; and the lag times in changing community relationships.

## Efficiency

**The programme as a whole cost over £72 million to deliver** between February 2014 and June 2019, meaning 90% of total budget was spent after 90% of the programme's lifetime. In 2019 EQUIP-T reached 5,192 primary schools with close to 3.2 million pupils enrolled. Of the programme support activity (PSA) spending, over 60% took place at LGA level as decentralised funds, and variation in unit costs across LGAs suggests that implementation was not consistent.

**The programme lacked a detailed TOC, and this was a missed opportunity for evidence-based design and reflection.** A detailed TOC would have allowed the MA to precisely recount each of the steps and assumptions, consider the strength of the underlying evidence, and weigh up the risks to success based on the design. There was also no regular critical reflection on a TOC to guide programme adjustments. A more thorough commitment to this exercise (up front and repeatedly) could have led the programme to eliminate some of its interventions which were never as likely or were not proving to be effective (such as community components), allowing reallocation of resources to more effective interventions. The programme was designed to address clear needs in these regions and did adapt to changes in context, as seen for example by the addition of an infrastructure component in 2017. Whilst the programme was intended to provide broad, systemic support, rather than being a narrowly targeted intervention, a greater focus on the evidence might have led the design to be less comprehensive and to concentrate on activities with greater chance of success.

**There were concerning limitations in the MA's budgeting, financial, and monitoring and evaluation (M&E) reporting systems.** Over the course of the programme, the MA moved from having comprehensive annual budgets broken down by meaningful categories, to having more of a rolling workplan and cash-budgeting approach. Furthermore, the coding system used in the budgets for decentralised funds did not correspond to the coding in the expenditure data. If budgets are not prepared and agreed over a meaningful timeframe (such as a year or a tranche), in the same categorisation system as expenditure tracking, then there is no framework to ensure that expenditure is guided by the objectives of the programme, or to assess whether expenditure and implementation is on track. Transparency around the budgets and expenditure tracking, at least to a wider annual review audience, diminished over the course of the programme. Furthermore, the difficulty in receiving a reliable dataset for LGA expenditure is not a sign of a well-functioning financial management system. In addition, the MA's M&E system displayed increasing weaknesses over time. The annual monitoring surveys contained appropriate quality control mechanisms, but the periodic Fact Sheets lacked quality control processes. The indicators in the logframe changed substantially year on year, compromising its ability to act as a monitoring tool beyond the very short term. The MA's Annual Reports became increasingly narrow, limiting their use as a monitoring tool to hold the programme to account.

## Sustainability

**Finally, many of the elements necessary for sustainability of parts of the EQUIP-T programme are in place, but not all.** Some central activities, such as teacher in-service training, training SCs to engage the community in school improvement, and holding DEMs, are seen as effective by stakeholders at the school and district level. Furthermore, Tanzania has the organisational and institutional capacity to continue these activities: the knowledge is already embedded in the government system. Sustaining these activities will need continued strong leadership and messaging from the top of government right through the levels of administration, with guidance and follow-up to make sure they become institutionalised. However, affordability is an issue, with all activities having at least an opportunity cost in terms of the time spent by public sector workers attending to these activities rather than other duties. In some cases, there are substantial budgetary costs too, as with holding training sessions. Many activities may appear 'zero-cost' but realistically have some minimal cost, and this is not currently affordable within LGAs' existing budgets.

## Recommendations

The recommendations below are intended to guide the Government of Tanzania and DFID in their adoption and adaptation of activities from EQUIP-T, as well as to identify broader areas where action would be valuable for improving equitable access to high-quality education at the primary level in Tanzania. The recommendations for government are particularly important given the government's commitment to sustain six elements of EQUIP-T: the in-service training model; School Readiness Programme (SRP); SIS; PTPs; IGA; and M&E. Meanwhile DFID is planning its next education support programme in Tanzania (called Shule Bora) now that EQUIP-T is coming to an end. Thus, the recommendations directly for DFID are broadly about the new programme design, but also may be relevant for future programmes. The recommendations for government are relevant for DFID too, in considering the details of the objectives and activities of the new programme. In addition, there are other programmes in the implementation stage, or being developed (such as the government's lead in shaping the new Global Partnership for Education Literacy and Numeracy Education Support (GPE LANES) 2 programme) which would benefit from these same recommendations.

## Recommendations for the Government of Tanzania

### Component 1: Improving teacher performance

1. **Under the National Framework for Teacher Continuing Professional Development (NF-TCPD), continue the in-service training model in EQUIP-T regions and scale this up to other regions that do not already have an effective school-based model.** The EQUIP-T in-service training model has been effective in improving teacher confidence and morale and in empowering teachers with a range of approaches which have contributed to improved learning outcomes. The model conforms to the principles and modalities set out in the NF-TCPD, and the numeracy, literacy, and other modules have been approved by the Tanzania Institute of Education (TIE).<sup>2</sup>

The model should continue at school level, which is lower cost, effective at providing ongoing support, and means teachers receive training despite high staff turnover. However, there does need to be intermittent face-time with experts (such as teacher training college (TTC) tutors or Teacher Resource Centre (TRC) staff) to refresh key ideas and troubleshoot issues. The revamping of TRCs, under the NF-TCPD, will help to support this need.

The school-based model should be institutionalised through formal recognition of the in-service coordinator role in schools (which may need promotion and additional salary), and by building this into the monitoring and quality assurance responsibilities of WEOs and school quality assurance (SQAs), as laid out in the NF-TCPD.

The participation and completion of in-service training should be built into the career progression framework to incentivise teachers to take part. The NF-TCPD states the intention to recognise and certify TCPD based on participation in standard modules that are accessible to all teachers.<sup>3</sup>

The best practices from the in-service training content should be integrated into pre-service training programmes.

At the time of this evaluation's fieldwork, it was clear that teachers faced challenges in managing large classes. The 'general effective pedagogy' module, rolled out by EQUIP-T after the qualitative fieldwork, will have been important in providing content on methods for large classes.

The government will need to increase budget allocation to education, including through LGAs, to sustain the in-service training model.

2. **Make use of the TLMs distributed under EQUIP-T.** The reading books are not currently being used, and they are a valuable resource. The government could include lessons on how to use reading books in both pre-service and in-service training, and include use of the books in curriculum guidelines (such as a minimum number of minutes per week for pupil reading practice).
3. **Review, with partners, the support for teachers on teaching children who do not speak Kiswahili as their mother tongue.** Many teachers are in contexts where many pupils do not speak Kiswahili as their mother tongue. This evaluation shows that these pupils are far behind their classmates in Kiswahili and maths, and wider evidence shows it could take many years for these pupils to catch up (Collier, 1989). Teachers would benefit from support in acquiring skills in approaches to teaching pupils in multilingual classrooms, and in putting these into practice. Such techniques could include games, group work, and use of translanguaging<sup>4</sup> (Heugh et al., 2019).

<sup>2</sup> The NF-CPD (p. 9) states that TCPD should be grounded in collaborative, inclusive, gender-responsive and participatory learning. It also emphasises the importance of non-residential models which allow teachers to be on task during school hours (p. 22). These elements are all central to EQUIP-T's model.

<sup>3</sup> See MoEST (2019).

<sup>4</sup> Translanguaging includes a range of processes in which bi-/multilingual people make use of the knowledge they have of many languages and how to use these languages. This can include alternating between two or more languages.

Interventions to help children catch-up before starting, such as the SRP, may also be beneficial, and this would be a useful focus for future research in Tanzania.

4. **Continue the focus on reducing classroom shortages and recruiting more teachers.** The evaluation confirms the extremely large pupil-to-classroom and pupil-to-teacher ratios and the challenges faced by teachers in this context. Building more infrastructure and recruiting more teachers are necessary to make these ratios more manageable.
5. **Review the teacher management policies which lead to high absenteeism and turnover.** The extremely high level of teacher absenteeism and turnover affects the usefulness of any intervention to improve teaching quality in the classroom. Improved monitoring and accountability for classroom attendance, and reducing the rate of transfers, would help address this. The recommendation to review transfer and turnover is also relevant to head teachers, WEOs, and LGA officers.

### Component 2: Strengthened SLM

6. **If the government is planning to continue using SIS, it needs improvements from the version seen by the evaluation team.** It should be reviewed in terms of ease of data entry for school staff and how to make the data useful at the school level. The SIS should not replace existing data collection systems until it is proven to work and all glitches have been removed for a whole annual cycle. More than one person per school should be trained in how to enter and use the data.

### Component 3: Strengthened district planning and management

7. **Continue and scale up DEMs across the country.** The introduction of DEMs has been successful at improving relationships between WEOs and the education department, sharing experience and learning, and allowing more efficient information sharing and management of schools. Ideally, DEMs should include an aspect of demand-driven training (akin to WEO continuous professional development – CPD). Based on the evaluation findings, this roll-out will need to involve some training, rather than just self-reading materials, with examples of best practice, continued leadership, and an emphasis from all levels of government to maintain and institutionalise the practices.
8. **Produce a standardised manual on WEOs' roles and roll these out through DEMs.** WEOs lack previous training on their roles and responsibilities, and appreciate the training they have received under EQUIP-T; this suggests the gap is likely to apply in the rest of the country and that a manual would be useful for new WEOs at least. After initial roll-out, the manual could be used within ongoing DEMs to discuss how WEO responsibilities should be performed and to guide the CPD aspect of these meetings.
9. **Continue providing responsibility allowances for WEOs and strengthen accountability for their performance.** The experience of the WEO grant under EQUIP-T has been that it unlocked more frequent visits to schools by WEOs. The government introduced a responsibility allowance in 2016 and although WEOs in EQUIP-T regions will now lose the value of the EQUIP-T grant, they will still have more resources than in the past, which is likely to make a substantial contribution to WEOs' performance. Although the use of the responsibility allowance is not stipulated (though perhaps should be), the government should enforce performance management of WEOs so that they are accountable for making visits to schools and supporting school improvement.

### Component 4: Stronger community participation and demand for accountability in education

10. **Review how to strengthen parental engagement and school accountability to the community, drawing on the best practices of SCs and PTPs.** SCs play an important and recognised role, including engaging with parents. The SC parent member is seen as a preferred

channel for parents to raise concerns, rather than via the PTP. In many cases, introducing PTPs has caused confusion and brought only minimal additional benefit in terms of wider parental engagement and school accountability. However, some PTPs are carrying out activities which are seen as beneficial within the school, even if they are not recognised as bringing the wider parental body closer to the school. It would be worthwhile considering options for encouraging the continuation and scale-up of these types of parent-led activities – involving PTPs and/or SCs. Future efforts to strengthen wider parental engagement and school accountability should recognise the central role of the SC, as well taking into account the constraints to participation and empowerment faced by parents, in deciding how to engage and represent them.

11. **Consider continuing distribution of school grants for IGA or school improvement.** The PTP grants and IGA grant have energised community mobilisation and participation, in terms of labour, in-kind resources, and interest in the school. The PTP grants allowed direct school improvement and gender welfare activities, whereas IGA projects, if successful, create an ongoing income stream for the school. The PTP grant does not require a PTP to be successful – the same aim could be achieved through the SC – and thus this recommendation can be considered separately. The cost analysis shows that the cost of these grants is actually rather small in comparison with current LGA budgets. Distributing grants for IGA or school improvement – in new regions which have not received them already – would help the SC to engage more with the school and the wider community.

### Sustainability of best practices

12. **Embed any new practices throughout the management chain.** Where activities are continued or scaled up to other regions, they need to be embedded in the responsibilities and monitoring systems throughout the chain: in how head teachers supervise teachers, how WEOs and SQAs supervise and help ensure quality in schools, how LGAs monitor WEOs, right up to regions and central ministries. At LGA level, DEMs are a platform to collect information on practices from the school and WEO level (such as in-service training). Regions should continue collecting performance and activity reports from LGAs as this accountability has contributed to an increased focus on results.
13. Recent trends in the education sector budget suggest that paying for the three activities studied here – teacher in-service training, PTP grants, and IGA grants with initial training – will not be affordable for the LGAs without additional financial support. The central government should take two actions if it wishes to continue replicating and scaling-up these activities across the country.
  - a. **Recognise the cost burden and provide a sufficient budget for these activities,** whether that budget is held and spent at national, regional, LGA, or school level. Government should not expect LGAs and schools to implement these activities if no provision is made for the costs. The government should also consider ways to reduce costs whilst maintaining satisfactory quality, as reviewed in the cost study (OPM 2020b).
  - b. **The Ministry of Education, Science and Technology (MoEST) and the President's Office Regional Administration and Local Government (PO-RALG) should strengthen the case for additional spending in the education sector,** to put to the Ministry of Finance and Planning (MoFP) and Parliament. This requires reviewing and assessing the evidence, and communicating this evidence to MoFP, the Cabinet, and parliamentarians, who each have a role in approving the final budget. Furthermore, MoFP should present annual budget allocation and disbursements at the joint annual sector review to provide additional accountability.

## Recommendations for DFID

1. **An evidence-informed TOC should be more explicit in the initial design of the programme, and it should be used for learning and adapting during implementation.** Setting out the assumptions more clearly at the start of the programme will allow more realistic reflection on the risks for programme success, and what needs to be monitored and tested before scaling. If DFID's new programme intends to 'test, adapt, and deliver at scale' then the learning process for adapting needs to be systematic, with critical reflection, adjustments, and documentation of those changes and the reasoning behind them. DFID should require this type of reflection and documentation as part of its Annual Reviews. The focus on evidence-based theories of change in the draft terms of reference (TOR) for Shule Bora is promising (DFID, 2019).
2. **Allow more room for discretion in the allocation of decentralised funds by the LGAs.** Whilst mitigating fiduciary risk should not be compromised, if LGAs had some room for decision making in how to use the funds (even within certain options/boundaries), this would increase local ownership, improve targeting towards local needs, and also allow LGAs to practice planning and budgeting skills.
3. **Programmes should have budgets for medium-term periods (such as annual) which are transparently reported against, and also robust financial tracking systems.** For future programmes, DFID should have closer sight in agreeing and monitoring progress against budgets for meaningful periods (such as annual) and component-level budgets. There should always be room for iterations as the context and programme change, but there should be agreed budgets to guide these iterations and explain changes. Budgets and expenditure need to be categorised in the same way to allow comparison. Expenditure tracking should be quality controlled and reliable. Expenditure and budget execution should be reported transparently in annual reviews.
4. **Ensure the financial system is set up to aid monitoring and accountability, and the assessment of value for money.** For example:
  - a. The activity coding structure should strictly relate to sub-component categories.
  - b. Put in place a level of classification in the financial data which relates to whether the activity was an overhead/development cost, or an implementation cost.
  - c. The category for implementation costs should further be coded to show which region each one is for, or whether it applies across all regions.
5. **Monitoring data should be comprehensive, regularly updated, and quality assured, and should ideally track actual beneficiaries.** Data should be collected and stored using a database. There should be documented standards and processes for ensuring data quality, and a sample of this data should be verified on at least an annual basis. Also at the design stage, there should be consideration of how the monitoring data can be made compatible with the coding of financial data. This monitoring system needs sufficient resourcing. The emphasis on bottom-up and comprehensive monitoring in the draft TOR for Shule Bora is positive.
6. **DFID should consider how to support the government in implementing the recommendations made above.** This would include integrating these recommendations into the design of DFID's new programme to support the government with the activities with highest impact. The programme could both sustain effective activities in the nine EQUIP-T regions, and incentivise activities for national scale-up through the results-based financing component.

## Lessons for future programmes in Tanzania and other countries

Various lessons of potential wider relevance for the design and implementation of education policies and programmes can be identified from the evaluation:

1. **Where teaching quality is low, and teachers lack access to regular training opportunities, the provision of teacher in-service training focusing on pedagogical practices is likely to improve learning outcomes.** However, the context is also important, and there will be more and less opportune times for such a training to be introduced if it is to be well-received. In the case of EQUIP-T, the change in Standards 1 and 2 curricula demanded new skills from teachers which made the training modules immediately relevant and useful for teachers.
2. **Where programmes have an objective to improve early grade teaching and learning, the existing language capabilities of pupils will be key to designing effective interventions.** Conducting baseline research to understand the language comprehension and pre-school experience would clarify whether interventions to increase school readiness and support transition to school, as well as support for teachers in these contexts, are necessary.
3. **As has been found in other studies (Glewwe and Muralidharan, 2015), the distribution of TLMs does not guarantee that they will be used.** In this case, most schools had received the supplementary reading books for early grade pupils but were not using them. Efforts to increase the use of TLMs will need to go further, to understand what makes teachers use them or not. This evaluation did not research why the materials were not being used, but factors which might be important for TLM distribution in general include: better sensitisation for teachers on what the TLMs are for and how to use them (integrated into the in-service training content); curriculum directives to use the TLMs; ensuring there are sufficient quantities for every child to be able to access the TLMs (so a slight shortage does not stop all children accessing the materials); reassuring teachers that they should be used and it is OK that the materials will get worn; and sensitising supervisors (head teachers, WEOs, and SQAs) on how to use the TLMs and on their roles in monitoring and supporting their use.
4. **The design of capacity building programmes in contexts with major system-wide constraints needs to take these into account and be cognisant of the risks to successful delivery and outcomes.** In particular, issues of high staff turnover, staff absenteeism, and staff shortages affect how the new capacity is put into use and retained in the system. As with EQUIP-T, refresher training is one way to mitigate these challenges, and programmes should consider whether more is needed.
5. **Increasing parental engagement and representation in school matters is very challenging, and is not ever likely to be quick or easy to change.** The cultural barriers and relationships are deeply engrained, and since parents are busy, there is a high opportunity cost to attending meetings, which cannot easily (and affordably) be overcome. Interventions targeting parental engagement need careful design and piloting, as well as setting realistic ambitions for the level of change possible in the programme period.
6. **This evaluation finds that continuing many of the activities will depend on strong leadership and availability of funds, which are by no means certain.** The challenge of sustaining donor-led initiatives is universal in cases where funding comes to an end and governments are unlikely to find new domestic sources of funds to fill the gap. Programmes should assess sustainability early and support efforts to lobby for more domestic funds, whilst at the same time being realistic about the likelihood of sustainability and the degree of sustainability acceptable to stakeholders.
7. **Programmes which aim to have a widespread effect on learning outcomes via government systems do not need to tackle all system constraints at once.** Allocating more time for evidence-based design, development, and scrutiny of the TOC and regular monitoring of the TOC, as well as piloting, would be helpful, rather than using resources to implement immediately at

scale. The information from monitoring and piloting should be used for learning, which may lead to stopping activities showing poor signs of effectiveness.<sup>5</sup>

8. **Programmes which are designed to be adaptive would benefit from an evaluation system which is tailored and responsive, in order to provide accountability.** An evaluation design which intends to quantify impact is ideally best suited to programmes which faithfully implement their original design. Furthermore, in adaptive programmes, the formal reporting for accountability purposes (such as through logframes) should have more emphasis on the learning process of the programme, rather than highly specific outputs, which are likely to change regularly.
9. **The measurement of value-for-money criteria is dependent on having a financial system set up to aid monitoring and accountability, as well as having reliable monitoring data which can be linked to financial inputs.** Budgets and expenditure should be categorised and coded so that the activity or sub-component is clear (with no overlaps), the geographical unit is clear, and the cost can be identified as a development/set-up cost or a recurring implementation cost.

---

<sup>5</sup> The draft TOR for Shule Bora has a promising focus on choosing the appropriate scale for implementation based on evidence, as well as monitoring and learning to adapt interventions.

## Table of contents

Acknowledgements	i
Executive summary	ii
List of tables, figures, and boxes	xvi
List of abbreviations	xviii
1 Introduction	1
1.1 Objectives of the IE	1
1.2 Overview of the evaluation process	1
1.3 Report structure	1
2 Programme design	3
2.1 Introduction to the programme	3
2.2 TOC	4
2.3 Implementation progress	9
2.4 Education sector context	10
3 Evaluation framework and methodology	12
3.1 IE design and methods	12
3.2 Endline design	13
4 Programme-level findings	19
4.1 To what extent was the programme appropriately designed to achieve its results?	19
4.2 What has been the impact of the programme on pupil learning outcomes?	20
4.3 Were targets for improved learning outcomes achieved? Did learning gaps narrow in favour of marginalised groups (girls, non-Kiswahili speakers)?	22
4.4 What factors explain these results?	23
4.5 To what extent was the programme cost-effective?	25
4.6 To what extent was the programme implemented in line with its agreed budget and timetable?	28
4.7 How well did key management arrangements support programme implementation?	30
4.8 Programme-level summary	33
5 The teacher component	34
5.1 Did the teacher component achieve its outcome targets for teacher performance?	35
5.2 What factors explain the results?	36
5.3 To what extent are factors favouring the sustainability of the teacher professional development model in place?	41
5.4 Teacher component summary	43
6 SLM component	45
6.1 Did the SLM component achieve its outcome targets for SLM?	45
6.2 What factors explain the results?	47
6.3 To what extent are factors favouring the sustainability of the SLM component in place?	50
6.4 SLM component summary	50

7	District component	51
7.1	Did the district component achieve its outcome targets for district planning and management?	52
7.2	What factors explain the results?	54
7.3	To what extent are factors favouring the sustainability of the district component in place?	59
7.4	District component summary	60
8	Community component	61
8.1	Did the community component achieve its outcome targets for community engagement?	63
8.2	What factors explain the results?	64
8.3	To what extent are factors favouring sustainability in place?	70
8.4	Community component summary	72
9	Conclusions, recommendations, and lessons	74
9.1	Conclusions	74
9.2	Recommendations	78
9.3	Lessons for future programmes in Tanzania and other countries	82
	References	84
	Annex A Agreed terms of reference	87
	Annex B Final Endline evaluation questions	93
	Annex C EQUIP-T Logframe, September 2019	100
	Annex E Accordance with international best practice	112
	Annex F Summary of PTP objectives and findings from fieldwork on PTP activities	116
	Annex G Recommendations for government: details	122

## List of tables, figures, and boxes

Table 1. EQUIP-T programme components and sub-components from mid-2017 .....	3
Table 2. Selected beneficiary and intervention data.....	9
Table 3. Summary of findings at programme level .....	33
Table 4. Summary of findings for the teacher component .....	43
Table 5. Summary of findings for the SLM component.....	50
Table 6. Summary of findings for the district component.....	60
Table 7. Summary of findings for the community component.....	72
Table 8. What the impact evaluation will measure .....	90
Table 9. Programme level evaluation questions.....	93
Table 10. Component 1 evaluation questions: Improved access to quality education .....	94
Table 11. Component 2 evaluation questions: Strengthened school leadership and management .....	95
Table 12. Component 3 evaluation questions: Strengthened district planning and management .....	96
Table 13. Component 4 evaluation questions: Stronger community participation and demand for accountability in education.....	97
Table 14. Stakeholder consultations and events—from dissemination of quantitative endline findings to plans for dissemination of final endline findings .....	107
Table 15. EQUIP-T Impact Evaluation Reference Group members (Invited January 2020) .....	109
Table 16. Endline impact evaluation team members and roles .....	110
Table 17. Mapping of findings of PTP activities against suggested activities .....	118
Table 18. Recommendations with responsible institutions and priority.....	122
Figure 1. EQUIP-T MA identified constraints on children’s capability to learn .....	5
Figure 2. EQUIP-T MA TOC in inception .....	6
Figure 3. EQUIP-T MA updated TOC for programme extension .....	7
Figure 4. High-level results chain from evaluation TOC .....	8
Figure 5. EQUIP-T programme districts (prior to the extension in 2017) and the IE district sample.....	12
Figure 6. Sequencing the endline .....	14
Figure 7. Trends in pupils’ Kiswahili skills and programme impact.....	21
Figure 8. Trends in pupils’ maths skills and programme impact.....	21
Figure 9. Total PSA expenditure, by category and component .....	26
Figure 10. LGA unit cost: average IGA grant per recipient school.....	27
Figure 11. LGA unit cost: average PTP grant .....	27

Figure 12. Main results chain for EQUIP-T sub-component 1A teacher performance .....34

Figure 13. Main results chain for EQUIP-T component 2, SLM.....45

Figure 14. Main results chain for EQUIP-T component 3.....51

Figure 15. Main results chain for EQUIP-T sub-component 4A community participation and  
accountability.....61

Figure 16. Main results chain for EQUIP-T sub-component 4B conducive learning environments .....62

Figure 17. Suggested activities for PTPs (without needing PTP grants) ..... 117

Box 1. Brief introduction to the endline technical studies ..... 14

## List of abbreviations

3Rs	Reading, writing, and arithmetic
ADEM	Agency for the Development of Education Management
CENA	Community Education Needs Assessment
COL	COL
CPD	Continuous Professional Development
DEM	District education meeting
DEO	District education officer
DFID	Department for International Development
EL Cost	Endline Cost Study
EL Qual	Endline Qualitative Report
EL Quant	Endline Quantitative Report
EPforR	Education Program for Results
EQUIP-T	Education Quality Improvement Programme in Tanzania
FGM	Female Genital Mutilation
GPE	Global Partnership for Education
IE	Impact evaluation
IGA	Income-generating activity
JUU	<i>Jiamini Uwezo Unao</i> ('Be confident, you have the ability, you can do it')
LANES	Literacy and Numeracy Education Support Programme
LGA	Local Government Authority
M&E	Monitoring and evaluation
MA	Managing Agent
MKUKUTA	<i>Mkakati wa Kukuza Uchumi na Kupunguza Umaskini Tanzania</i> (National Strategy for Growth and Reduction of Poverty)
MoEST	Ministry of Education, Science and Technology
MoFP	Ministry of Finance and Planning
NECTA	National Examinations Council of Tanzania

NF-TCPD	National Framework for Teacher Continuing Professional Development
OC	Other charges
OECD-DAC	Organisation for Economic Co-operation and Development – Development Assistance Committee
OPM	Oxford Policy Management
PEDP III	Primary Education Development Programme III
PFM	Public financial management
PO-RALG	President’s Office Regional Administration and Local Government
PSA	Programme support activity
PTP	Parent–teacher partnership
SBCC	Social and Behaviour Change Communication
SC	School Committee
SDP	School Development Plan
SIS	School information system
SLM	School leadership and management
SPMM	School performance management meeting
SRP	School Readiness Programme
SQA	School quality assurance/assurer
TLMs	Teaching and learning materials
TIE	Tanzania Institute of Education
TOC	Theory of change
TOR	Terms of reference
TRC	Teacher Resource Centre
TTC	Teacher training college
TZS	Tanzanian shilling
USAID	US Agency for International Development
WEO	Ward Education Officer

# 1 Introduction

## 1.1 Objectives of the IE

This report is the final endline report of the IE of EQUIP-T, a Government of Tanzania programme funded by DFID.

The overall objectives of the EQUIP-T IE 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 IE to policymakers and key education stakeholders.

This report has two primary audiences: the Government of Tanzania, which is considering how to adopt and adapt activities from EQUIP-T and scale up nationally; and DFID, for accountability and learning, particularly as it considers future education programming in Tanzania. As the EQUIP-T programme started winding down in mid-2019, with final closure in January 2020, these findings will not inform future implementation of EQUIP-T, but of course are relevant to future efforts which build on or learn from EQUIP-T's experience. More widely, the report is intended to be informative for a variety of education sector stakeholders, such as other development partners, non-governmental organisations involved in the education sector, research organisations, and academics. Further afield, the findings are also likely to be of interest to an international audience wanting to learn about education programmes which deliver impact on learning outcomes in low-income settings.

For readers wishing to read only a selection of the evaluation, the Executive Summary is the most important section. This is intended to be a succinct overview of the objectives, methodology, conclusions, recommendations, and lessons for future programmes elsewhere.

## 1.2 Overview of the evaluation process

The endline forms part of a mixed-methods evaluation that began with a baseline in 2014, was followed by a midline in 2016, and ends with this report in 2020. The endline is made up of four products: a quantitative endline study carried out in 2018 (OPM, 2019a); a qualitative study conducted in 2019 (OPM, 2020a); a cost study, also conducted in 2019 (OPM, 2020b); and this final summary report, which draws on all three technical studies.

The programme has 10 sub-components, of which eight are at least partially included in the scope of the evaluation and were the focus of primary quantitative or qualitative data collection. Analysis of programme costs covers the full scope of the programme, and narrows down on three particular activities for assessing the fiscal affordability of scaling these up nationally for the Government of Tanzania.

## 1.3 Report structure

The report is structured as follows: Chapter 2 introduces the programme's design, its components, the evolution of its TOC, implementation progress, and the education sector policy context in Tanzania. Chapter 3 sets out the overall IE design and methods, and how the endline technical studies have been drawn on for this report. The findings are set out in Chapters 4 to 8, which in turn cover the

programme-level findings, and the teacher, school leadership, district planning, and community-level components of EQUIP-T. Chapter 9 covers the evaluation's final conclusions, recommendations, and lessons.

## 2 Programme design

### 2.1 Introduction to the programme

EQUIP-T began in 2014 as a four-year, Government of Tanzania programme funded by DFID. The aim of the programme has been to increase the quality of primary education and improve pupil learning outcomes, in particular for girls. The programme planned to reach 2.3 million primary pupils, about one-quarter of total primary enrolment in Tanzania. Over time, the programme has been extended in terms of duration, geography, and activities. Now scheduled to finish in January 2020, the programme was extended from the original five to seven (in 2015), and eventually nine (from 2017), of the most educationally disadvantaged regions in Tanzania.<sup>6</sup> The budget was extended in 2017 from approximately £50 million to £90 million. The programme has an MA, Cambridge Education, which works with the government to deliver the programme. The MA is responsible for £80 million of the total EQUIP-T budget, and the activities overseen by the MA are the focus of the IE.

EQUIP-T comprises five components (and 10 sub-components following the extension), as in Table 1.

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

Component and sub-components <sup>1</sup>	Scope of IE
<b>1 Improved access to high-quality education</b>	
<b>1A Improving teacher performance</b>	✓
<b>1B SRP</b> and satellite schools (non-construction)	X <sup>2</sup>
1C Classroom and satellite school construction	X
<b>2 Strengthened SLM</b>	
<b>2A SLM capacity building</b>	✓
2B SIS	✓
<b>3. Strengthened district planning and management</b>	
<b>3A District capacity building</b>	✓
3B LGA (district) grant monitoring	✓
<b>4 Stronger community participation and demand for accountability in education</b>	
<b>4A Community participation and accountability</b>	✓
4B Conducive learning environment for marginalised children, particularly for girls and children with disabilities	✓
<b>5 Improved learning and dissemination</b>	
<b>5 Learning and dissemination</b>	X

Source: OPM 2018, p. 13. Notes: (1) The original components, planned since inception, are in bold. The rest are new components introduced in the 2017 extension. (2) The midline qualitative study did collect some information on the SRP, but its coverage is very limited.

The programme started with just five of the sub-components (highlighted in bold above), which 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. Overall, the emphasis of this first set of EQUIP-T interventions is on strengthening the education system to deliver

<sup>6</sup> There are 26 regions in mainland Tanzania. The original regions in the EQUIP-T programme are Dodoma, Kigoma, Tabora, Shinyanga, and Simiyu. These were followed by Lindi and Mara, and later Singida and Katavi.

**high-quality** education, and these are the focus of the IE because these initial interventions have been tracked since baseline. (See OPM 2019a for more detail on the extension and components.)

At the time of the extension design in 2017, EQUIP-T discovered that **access** to education, particularly for remote and marginalised children, is a more serious problem than had initially been understood. The problem 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, 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. To support national adoption and the scale-up of successful parts of the programme, EQUIP-T has an institutional strengthening and sustainability strategy integrated into its TOC.<sup>7</sup>

It is important to highlight that EQUIP-T introduced a decentralised funding mechanism for its programme support funds (approximate budget of £37 million) in the 2015/16 financial year. Activities under all five components, but particularly components 1 to 4 (excluding centrally procured materials and contracts and region-level training), are funded by districts. DFID's programme funds are channelled and managed through the Government of Tanzania's PFM system to the districts (also known as LGAs).

The original scope of the IE included four of the original five components: teacher performance (1A); SLM (2A); district planning and management (3A); and community participation and accountability (4A). Learning and dissemination (5) was the programme's internal M&E and was not within the scope of the IE, meaning that the evaluation did not examine these activities in detail or study their pathway to impact. An expanded scope for the IE was agreed in the Endline Planning Report Part I (OPM, 2018), covering some, but not all, of the new sub-components. 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. Construction was not included because of the very different nature of the intervention and the timelines anticipated to lead to outputs and eventually improved learning outcomes. However, the evaluation's estimate of the impact of EQUIP-T on pupil learning picks up the combined effect of all interventions, so in this sense any impact due to construction is included. The other new sub-components are partly included, which means that the evaluation produces some (quantitative or qualitative) evidence related to these sub-components.

The next section explains the TOC. It is followed by a brief summary of the interventions and implementation progress.

## 2.2 TOC

With adaptations to the programme's design and interventions, the TOC has evolved over time. The evaluation uses a TOC developed specifically for the evaluation, with inputs from the MA. These different versions are explained in this section.

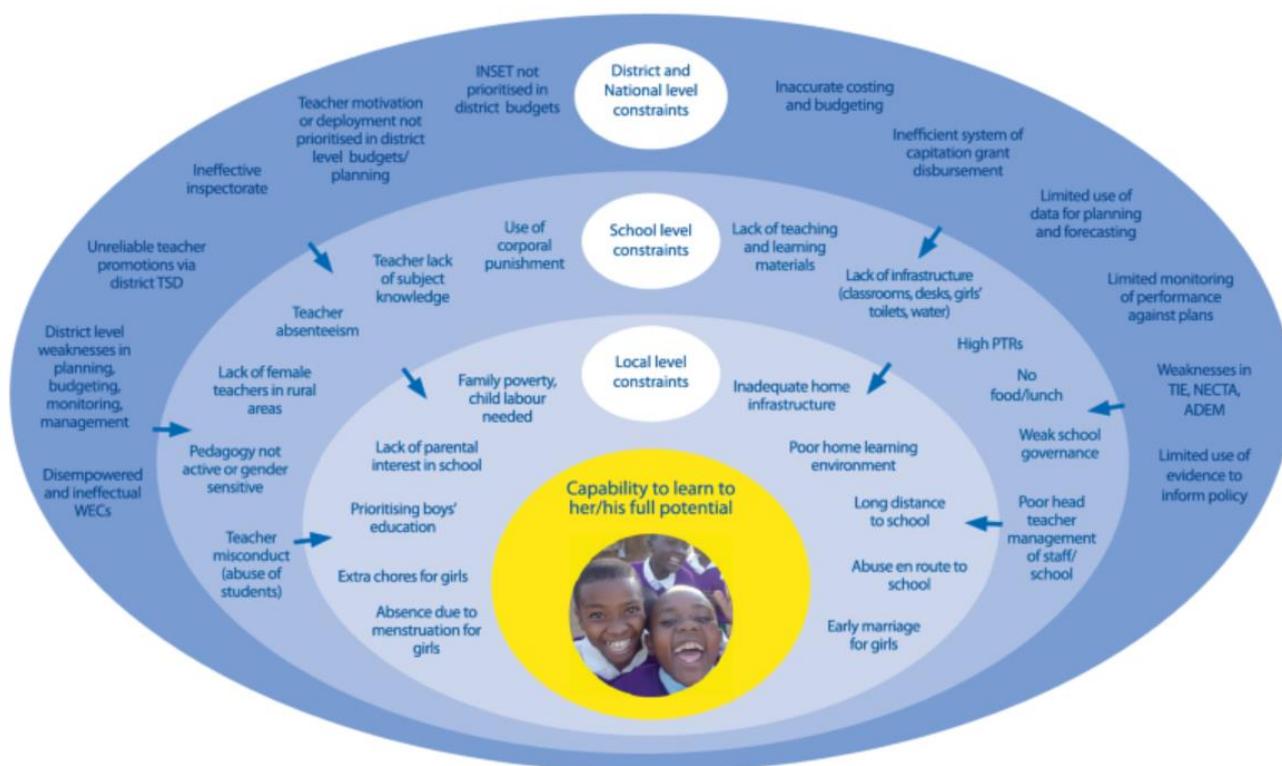
### 2.2.1 The EQUIP-T MA TOC

The programme began in its inception phase with an overarching TOC which conceptualised the components as mutually reinforcing in overcoming the barriers identified to pupils' learning at local, school, and district/national levels. The constraints identified as hindering children's ability to learn

<sup>7</sup> This evaluation does not directly evaluate the institutional strengthening and sustainability strategy; however, one of the evaluation criteria used here is whether the conditions for the sustainability of key parts of the programme are in place.

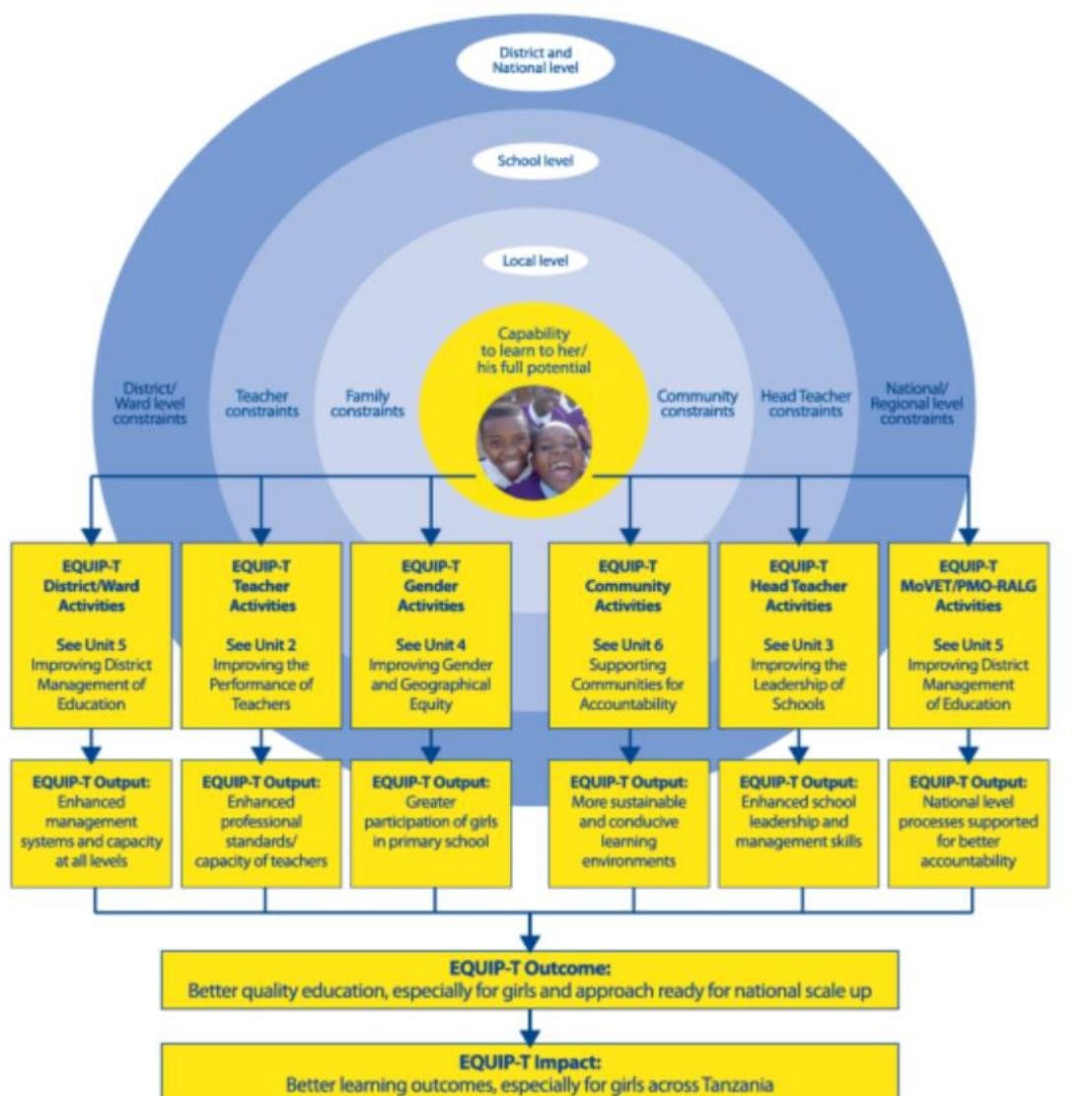
were set out as a diagram, shown in Figure 1. Taken together, the programme’s components were expected to lead to better quality education, especially for girls (EQUIP-T outcome), and to improved learning outcomes, especially for girls across Tanzania (EQUIP-T impact). The TOC was set out as a diagram with general statements of change: the interventions aimed to overcome the identified constraints and so lead to the outputs, outcome, and impact (see the EQUIP-T MA inception report, 2014). No further narrative was provided on the TOC, of the links in the chain, the assumptions and risks, or the underlying evidence for this design. The diagrammatic TOC is shown in Figure 2.

**Figure 1. EQUIP-T MA identified constraints on children’s capability to learn**



Source: EQUIP-T MA (2014)

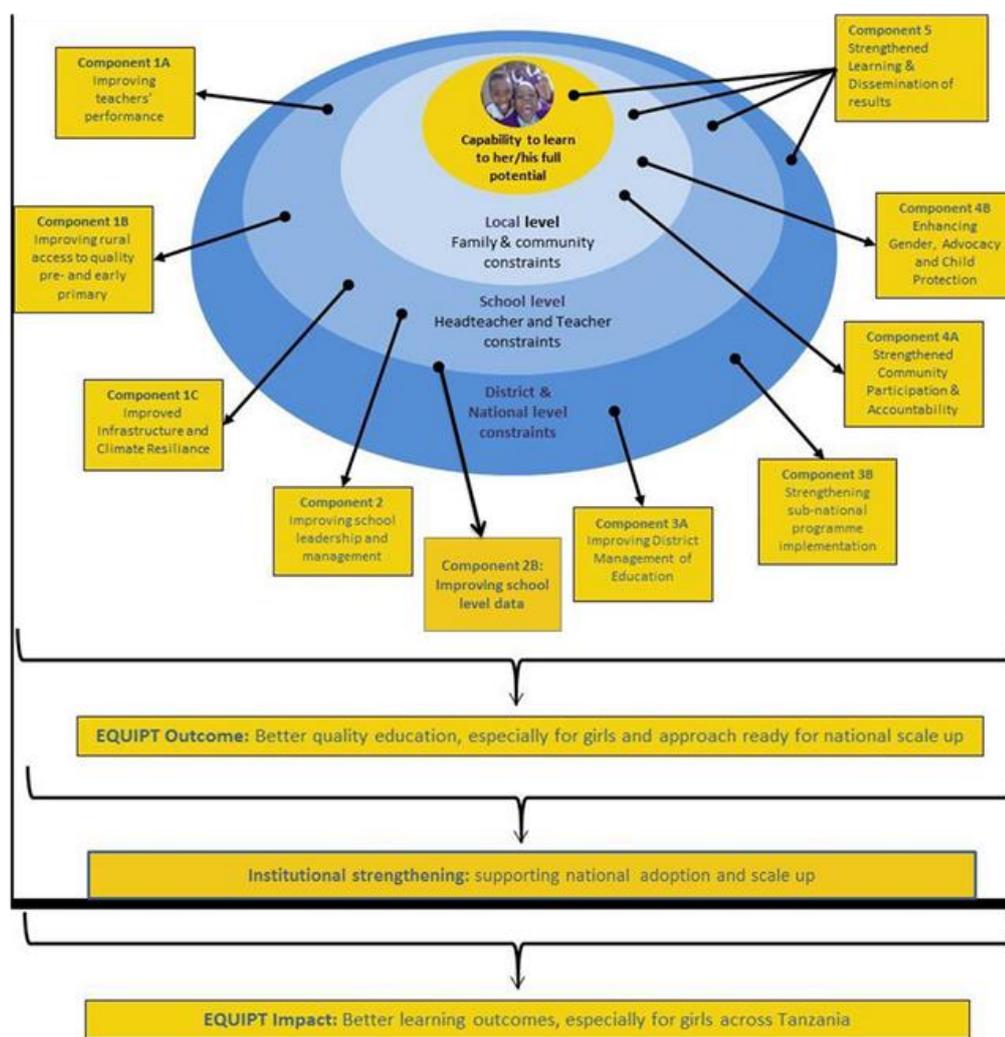
Figure 2. EQUIP-T MA TOC in inception



Source: EQUIP-T MA (2014)

As mentioned, with the extension design, EQUIP-T identified **access** to education as an additional constraint to children’s education and introduced new sub-components. At this point, the MA revised its over-arching TOC, as shown in Figure 3. **Again, this diagram shows the general direction of change but does not include the detailed causal pathways or the assumptions within the TOC. It is considered standard practice for a TOC to include detailed causal pathways and assumptions.** The EQUIP-T MA’s updated TOC from the extension phase is shown below.

Figure 3. EQUIP-T MA updated TOC for programme extension



Source: EQUIP-T MA (2017a)

### 2.2.2 The TOC used in the evaluation

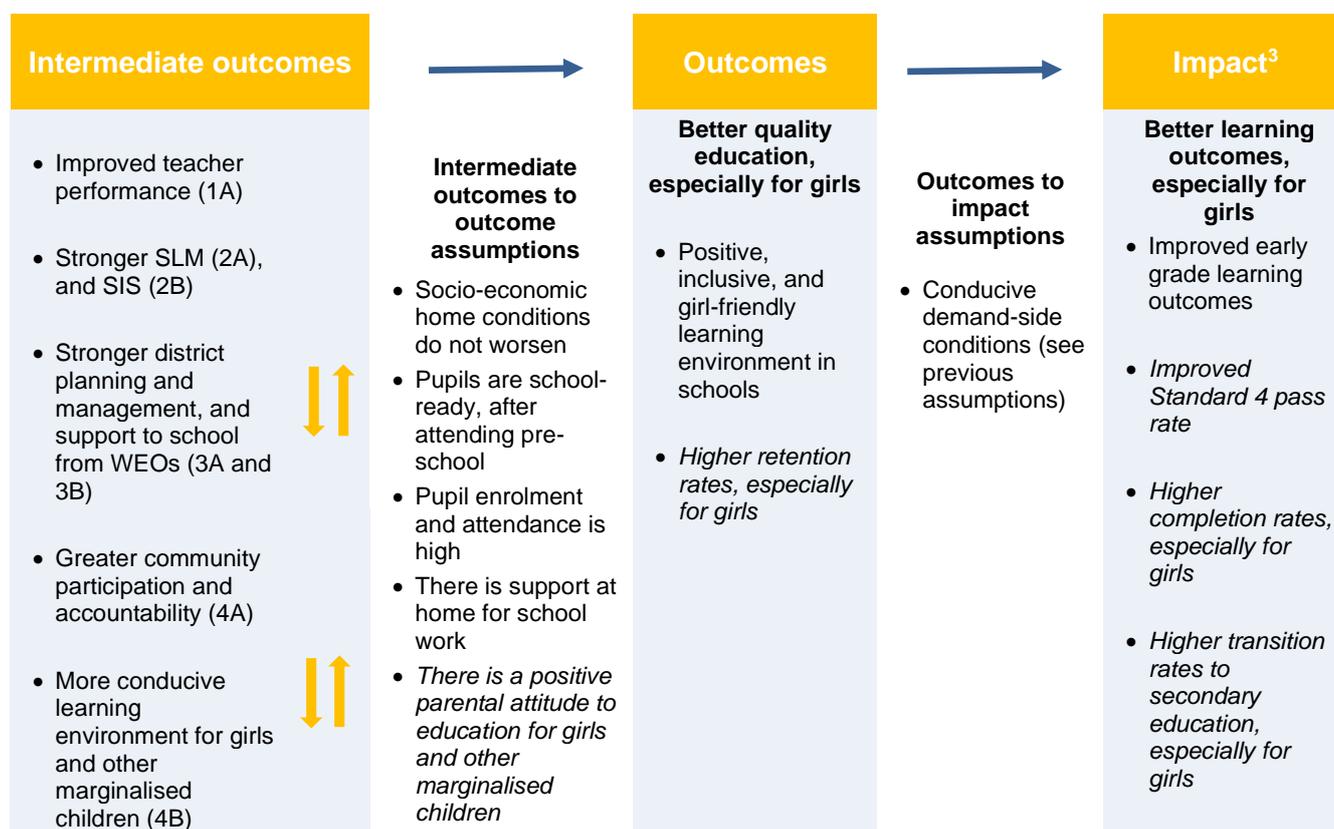
At the baseline, OPM was asked by DFID to expand and assess the TOC. This process began by explicitly mapping out the key causal pathways through which the programme expected to see change. The next step was to assess the strength of the key assumptions underpinning each link, based on the wider literature, as well as contextual information from the baseline data. This analysis was shared with the MA for comment, and can be found in the baseline evaluation report (OPM, 2015a).

By the time of the midline study, the detail of the activities being implemented by the programme had changed since baseline, and due to the extension in 2017, the programme design changed again between midline and endline. Thus, at both midline and endline, the evaluation developed more

detailed versions of the TOC for the sub-components in scope.<sup>8</sup> These were developed through interviews with EQUIP-T programme staff, and a joint workshop in early 2018 at which the evaluation team and EQUIP-T staff documented the detailed results chains<sup>9</sup> implicit in the overarching TOC, including the key assumptions underpinning the links.<sup>10</sup>

In the evaluation’s endline TOC, each of the four components (and their sub-components) leads to the intermediate outcome indicated in the name of the component (for example, improved teacher performance). The highest level of the TOC shows how these intermediate outcomes lead through to the ultimate outcomes and impact of the programme. Figure 4 illustrates these causal links as elucidated by programme staff, as well as the assumptions that must hold for these chains to work. Critically, there is a synergistic aspect to the intermediate outcomes at this level, in that in theory they contribute to each other. For example, stronger district management should lead to better support and supervision of schools by WEOs, which in turn should contribute to both better SLM by head teachers and more accountability and higher performance by teachers. Similarly, greater community participation and accountability should contribute to head teachers managing their schools more effectively, and to teachers being in their classrooms and focusing on teaching more.

**Figure 4. High-level results chain from evaluation TOC**



Source: OPM (2018). Notes: (1) This is a partial high-level results chain, which covers programme components included in the endline evaluation. It excludes the institutional strengthening aspect of the TOC, whereby the programme seeks to support the national adoption of successful aspects of the EQUIP-T programme. (2) The indicators and assumptions in italics are not being measured in the IE. (3) The impacts here were identified by the MA but do include changes which would

<sup>8</sup> This means that some inputs and activities took place that were not in scope and are not covered here, especially activities conducted after the quantitative endline and not in scope of the qualitative endline, such as School Report Cards.

<sup>9</sup> A results chain is the path from the receipt of inputs to the delivery of outputs, and contribution to intermediate outcomes and impact.

<sup>10</sup> This was not exhaustive, covering all results-chains embedded in the TOC, but instead focused on the main interventions within the scope of the evaluation, expected changes to follow from these, and key assumptions.

normally be defined as outcomes (e.g. completion rates) because they lead to an ultimate desirable goal, such as a more productive and employable workforce.

The high-level results chain is relevant to the programme level findings in Chapter 4. The results chains for the components within scope of the evaluation are included in Chapters 5 to 8.

## 2.3 Implementation progress

The introductions of Chapters 5 to 8 give an overview of the inputs provided under components 1 to 4. However, a brief (and by no means comprehensive) introduction to the inputs is as follows:

- Component 1A: in-service training for teachers focusing on numeracy, literacy, and gender-responsive pedagogy (and general effective pedagogy carried out after evaluation data collection), supported by videos on the training modules; distribution of TLMs for literacy and numeracy.<sup>11</sup>
- Component 2: training for head teachers on SLM, setting up SDPs, and use of the SIS, as well as the distribution of tablets for the SIS to schools. The SIS is a management tool used for education planning, monitoring, and decision making, based on data collected at school level.
- Component 3: training for LGA staff on planning and budgeting; provision of grants for implementing activities across all components, and training and mentoring on using the grants; and establishment of DEMs. WEOs were provided with motorbikes and grants, and training on their roles and responsibilities.
- Component 4: training for SCs and establishment of PTPs; provision of PTP grants and IGA grants to establish IGA projects; provision of school noticeboards; conducting community education needs assessments (CENAs); establishment of school clubs.

More detail on implementation progress is given in OPM (2019a) Annex B and OPM (2020a) Annex C.

To give a sense of the scale of the programme, Table 2 below includes a selection of indicators on the numbers of beneficiaries and the distribution of inputs across the components.

**Table 2. Selected beneficiary and intervention data**

	Original seven regions	Two extension regions 2017	All nine regions	Notes
<b>Basic programme data</b>				
Number of primary schools	4,486	706	5,192	In 2019
Number of pupils enrolled	2,728,246	427,543	3,155,789	In 2019
Number of teachers	48,923	7,216	56,139	In 2019
Number of LGAs	51	12	63	In 2019
Number of wards	1,130	190	1,320	In 2019
<b>Component 1 Improved access to high-quality education</b>				
Teachers training colleges working with EQUIP-T	14	3	17	
Teachers trained in Literacy Modules 9–13	12,574	3,636	16,210	
Supplementary Readers Distributed	286,000	(1)	286,000	DM for Mara

<sup>11</sup> Component 1B involved the establishment of the SRP – a 12–16 week pre-school programme for children who otherwise would likely not attend pre-school, delivered by community teaching assistants. Component 1C included the community-led construction of up to 220 satellite schools and 230 finished classrooms, including roofing.

'Teacher Read Aloud's Distributed	41,116	(1)	41,116	DM for Mara
Big Books Distributed	4,386	3,864	8,250	
School Readiness Centres	2,737	441	3,178	No date
Children enrolled in the SRP	121,547	21,630	143,177	In 2017
Satellite schools to be constructed	102	24	126	
Unfinished classrooms to be completed	102	24	126	
<b>Component 2 Strengthened SLM</b>				
Heads, assistant heads, and WEOs trained on SLM module 1	7,890	0	7,890	
Heads trained on SIS and given tablets	4,454	706	5,160	
<b>Component 3 Strengthened district planning and management</b>				
Regional and LGA staff trained on Decentralised Planning and Budgeting	287	DM		
Motorbikes distributed	1,009	(2)	1,009	
<b>Component 4 Stronger community participation and demand for accountability in education</b>				
Noticeboards distributed	4,514	704	5,218	
PTPs established	4,438	706	5,144	
PTP establishment manual distributed	29,348	1030	30,378	
Schools receiving IGA grants	2,236	0	2,236	
Sources: EQUIP-T MA Monitoring data (Summary master data revised 25 February 2019) and data sent by MA as feedback on draft report. Notes: (1) Distributed by TIE through GPE LANES programme. (2) Distributed by GPE LANES programme. DM = data missing. Total for all nine regions is left blank where data for the two extension regions is missing.				

## 2.4 Education sector context

### 2.4.1 Context at the start of the programme

At the time of EQUIP-T's inception in 2013, the education sector was guided in turn by the Education and Training Policy (1995), the Education Sector Development Programme (2008–2017), and the third Primary Education Development Programme (PEDP III, 2012–2016). These sought to achieve Tanzania's Vision 2025 (to achieve middle-income status by 2025), and the objectives in the National Poverty Reduction and Development Plan (MKUKUTA II 2010–2015). EQUIP-T was designed in line with PEDP III.

As with many other countries, Tanzania had seen notable successes in increasing access to education, but without commensurate progress in learning outcomes. At the time of EQUIP-T's inception, independent national learning assessments were showing a learning crisis in the nation's schools: in 2012, only one quarter of Standard 3 pupils had Standard 2 level literacy in Kiswahili, and only four in 10 pupils had Standard 2 numeracy skills (Uwezo, 2013). A number of issues contributed to this; for example, a 2014 survey estimated that children were receiving less than half of the intended hours of teaching time, due to teachers' absenteeism from school and engagement in non-teaching activities (World Bank, 2015).

EQUIP-T was designed to target some of the lower performing districts in the country, and thus results in terms of education would be even worse in the programme regions than nationally. The evaluation's baseline in 2014 found significant differences between the initial five regions in the programme and the national picture: these regions were more rural, had higher pupil–teacher ratios, lower school enrolment rates, and higher indicators of poverty and disadvantage (OPM, 2015a). In terms of

learning, pupils in the initial five regions on average performed much worse than the national average on the primary school leaving examination. Specifically, close to half of schools in programme districts did not even manage a 25% pass rate in this examination, whilst only one fifth of schools nationally were in this situation (OPM, 2015a).

## 2.4.2 Changes in the policy context

The quantitative endline report summarises the major national policy changes and other large education programmes which have affected primary education across the country during the period since the start of EQUIP-T in 2014. These have not changed since this report. In brief, they are:

- **Fee-free basic education policy** which states that no fees and or other compulsory contributions would be charged for pupils from Standard 1 to Form 4, announced in December 2015;
- **School capitation grants** paid monthly and directly into schools' bank accounts, from January 2016;
- **New competency-based early grade curriculum** which focuses on reading, writing, and arithmetic (3Rs), rather than a larger set of subjects, introduced in mid-2015; and
- **Human resource management** in which government introduced higher minimum qualifications for head teachers (diplomas) and WEOs (degrees) than previously, as part of the government's professionalisation policy in 2017. This means that under-qualified head teachers and WEOs had to leave their positions. At the same time a large number of civil servants, including teachers, were dismissed 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, with the aim of improving public services.

Three major primary education development programmes have been implemented alongside EQUIP-T: GPE LANES, Education Program for Results (EPforR), and Tusome Pamoja.<sup>12</sup> The EQUIP-T MA worked closely with these programmes for shared learning and combined efforts. Overall, the evaluation assesses that these programmes will have had very minimal contamination<sup>13</sup> on the quantitative impact estimates (more details are given in OPM, 2019c – Volume II of the Quantitative endline, Chapter 3 Section 3.5.1).

Each of these policies and programmes are explained in more depth in OPM (2019a) Section 1.4.

---

<sup>12</sup> LANES is funded by the Global Partnership for Education, and is largely national in coverage. EPforR is funded by the World Bank, Swedish International Development Cooperation Agency, DFID, and the Government of Tanzania, and is largely national in coverage. Tusome Pamoja is the USAID-supported programme in four mainland regions.

<sup>13</sup> Contamination occurs when members of treatment and/or comparison groups have access to another intervention which also affects the outcome of interest.



This analysis produces estimates of the impact of the EQUIP-T programme **as a whole** on impact-level indicators (pupil learning achievement), but is not able to quantify the impact of the different EQUIP-T components separately on these indicators. Also, impact cannot be estimated separately for boys and girls, though the evaluation does assess descriptive trends by gender.

**2) Rigorous factual analysis to explain programme impact:**<sup>14</sup> This approach combines quantitative survey data on trends in programme areas with qualitative research findings, together with other secondary sources, in order to establish key channels of programme influence, or the reasons for lack of change, addressing research questions structured around the programme's TOC (the evaluation's version). In other words, this analysis seeks to understand whether, how, and why changes happened as anticipated in the TOC or not, and whether there are any important unanticipated changes (positive or negative).<sup>15</sup>

The baseline and midline qualitative studies explored perceptions of change and explanations from stakeholders from school and community up to national implementation level, looking broadly across the four components in the IE scope. At endline, the qualitative research provided greater depth into three focus areas. It allowed deeper analysis of the community and teacher in-service training sub-components, with the aim of exploring and explaining the quantitative findings. It also covered the district component, which had very limited coverage in the quantitative study and underpins the institutional sustainability of other activities implemented at decentralised level. The endline research questions were selected following consultation with the IE Reference Group and DFID, and the qualitative study used purposive sampling of 'best practice' cases to clarify what success looks like and why these cases were successful. More information is included in OPM (2020a), Section 3.

In addition, the cost studies carried out at midline and endline provide greater breadth to the evaluation by responding to questions on the efficiency and sustainability of the programme. The endline cost study uses data from the MA on budgets and expenditure, and monitoring data on beneficiaries to analyse the cost drivers and efficiency of various interventions. The study also collected primary data on unit costs and input quantities in order to make simple projections<sup>16</sup> of the costs and affordability of rolling out three activities nationally – teacher in-service training, PTPs, and IGA. More information on the methodology is included in OPM (2020b), Section 2.

## 3.2 Endline design

### 3.2.1 Sequencing of mixed methods

The EQUIP-T extension in 2017 led to a decision to split the endline evaluation into two parts, with the first (quantitative) part carried out in 2018, and the second (the qualitative and cost studies) in 2019. This was because the extension in 2017 included roll-out to two new regions, one of which was part of the control group for the quantitative impact estimates. The quantitative survey needed to be carried out as early as possible in order to avoid contamination of the impact estimates, due to any effects in the control districts. Delaying the qualitative work and cost study allowed for more time for implementation before collecting data. In the end, the design of the endline qualitative research

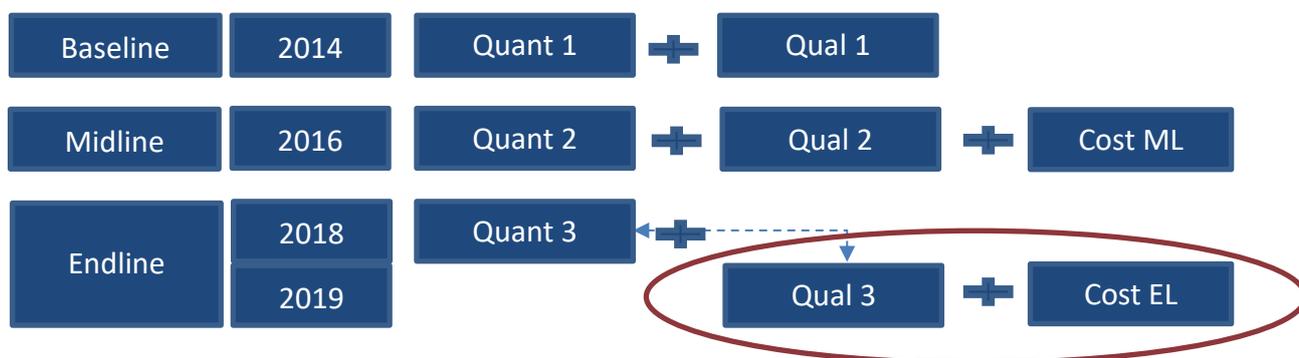
<sup>14</sup> The term 'rigorous factual analysis' comes from White (2009).

<sup>15</sup> The evaluation does not constitute a full theory-based evaluation (to the extent set out in White, 2009), mainly because it does not exhaustively cover all parts (all linkages) of the TOC. This is acknowledged in the agreed TOR.

<sup>16</sup> The analysis of costs of continuing and scaling-up uses the 'ingredients approach', as set out in Levin and McEwan (2001).

followed a few months after findings from the quantitative were analysed and disseminated. The sequencing is shown in Figure 6.

**Figure 6. Sequencing the endline**



The quantitative part of the endline evaluation focused on measuring impact on pupil learning, and on providing quantitative evidence on the effectiveness of the school- and community-level interventions up to early 2018. These results then fed into the design of the qualitative and cost endline research in 2019, and helped to prioritise research themes for follow-up (continuation of evaluation narrative).

The endline studies combine various elements of mixed-methods approaches.<sup>17</sup> The sequencing allowed the findings from the quantitative endline to be used to select the sampling and shape the research questions for the qualitative and cost studies. The qualitative and cost studies are complementary in that they elaborate how and why change has or has not taken place. They also allow for greater breadth and scope in the overall evaluation by assessing components and processes not suitable for the quantitative survey.

The scope of all three endline technical studies was discussed with the IE’s Reference Group, and set out in Endline Planning Reports Part I and Part II (OPM 2018 and 2019b).

### 3.2.2 Endline technical studies

There are thus three technical studies which form the basis of the evidence used in this final endline report. Each of these studies contains full details of their methodologies: the sampling, instruments, and respondents; secondary data sources and analysis methods; and quality assurance processes. The final endline report also draws on the baseline and midline evaluation reports.

An introduction to the three technical studies is given in Box 1.

#### Box 1. Brief introduction to the endline technical studies

Endline Study	Methodology
Quantitative Report (OPM 2019a)	The quantitative report is based on the 2018 quantitative survey. The sample is a panel of 200 schools: 100 programme treatment schools, and 100 control schools acting as a counterfactual. The study estimates quantitative attributable impact on a small number of indicators, including pupil learning, and collects data on trends in programme schools of input, output, and outcome level indicators and assumptions. The programme treatment schools are representative of 17 districts in the original five regions of the programme,

<sup>17</sup> These elements are termed ‘development’, ‘complementary’, and ‘expansory’ in mixed-methods design (Greene, Caracelli, and Graham, 1989).

	which are broadly similar to the other 12 districts in the original five regions (see OPM 2019a, Section 2.4) but are not representative of all programme schools in all nine regions.
Qualitative Report (OPM 2020a)	<p>The qualitative report is itself is made up of two studies:</p> <p>The school/community study focused on the community component (PTPs, SCs, and IGA in particular) and the teacher in-service training. Data was collected from six schools/communities, purposively sampled as representing successful cases of active PTPs, with continuity in the head teacher and teachers. Respondents included head teachers and teachers, parent and teacher members of the PTP, and other parents, SC members, and community leaders.</p> <p>The district-level study focused on education planning and management at the district office level, as well as the support given by WEOs to schools. Data was collected from three districts, two sampled as having strong WEO support to schools, and one selected for being strong at district planning and management. The respondents were LGA officers from the education, the planning and treasury departments, and WEOs.</p> <p>Research instruments included in-depth interviews, focus group discussions, and participatory techniques such as community mapping.</p>
Cost Study (OPM 2020b)	The cost study used data from the EQUIP-T MA to analyse the costs and spending of the programme at MA and LGA level. It also estimated the costs of continuing and scaling up three activities of EQUIP-T (teacher in-service training, PTPs, and IGA) to all regions in the country. These three activities were chosen because (as with the focus of the endline qualitative study) there was government and DFID interest, and the costs will mainly be incurred at school and district levels.
<b>Source: OPM 2019a, 2020a, and 2020b</b>	

All publicly available technical studies, and the baseline and midline reports, are available here: [www.opml.co.uk/projects/assessing-equip-t](http://www.opml.co.uk/projects/assessing-equip-t).

In addition, the team developed a Management Background Note (OPM 2020c) for this final endline report, as an internal note for DFID. In the process of designing this final report, the evaluation team identified evaluation questions relating to the programme's management, under the broad criteria of efficiency. The scope of the EQUIP-T IE did not include an explicit evaluation of the way in which the programme was managed. The emphasis of the evaluation was on assessing whether the programme was successful in meeting its outcome and impact objectives (effectiveness), and then seeking to determine the extent to which different parts of the programme contributed to the measured impact. It did not include a process evaluation, which has its focus on the details of programme implementation and management. That said, during the course of the six years of the IE, the evaluation team has regularly reviewed core management documents and datasets – especially the Annual Reports, various editions of the logframe, and Fact Sheets (data on beneficiaries and programme outputs). In preparation for each round of IE research, in 2014, 2016, 2018, and 2019, the evaluation team also spent several days interviewing staff from the EQUIP-T MA – including the senior management team and all of the staff leading the programme's components. From this evidence base, the team developed the Management Background Note that highlights observations on the management of EQUIP-T, and this is used in the programme level findings in Section 4.7. The draft management note was also shared with the MA, who sent comments to the evaluation team and to DFID.

### 3.2.3 International best practice

As with all OPM research, the IE was guided by best practice in terms of ethical standards, safeguarding, and confidentiality. It was also implemented in accordance with the Paris Declaration Principles. Annex E sets out further details on the conformity with these best practices.

### 3.2.4 Final endline report design

Each of the technical studies until now has sought to answer a selection of evaluation and research questions which were defined in the planning stages of those studies, and together contribute to the IE's overall objectives, as set out in Section 1.1. For the final endline report, the evaluation team developed a series of evaluation questions to cover the entirety of the endline evaluation. The evaluation questions were drawn from the OECD DAC evaluation criteria: relevance, impact, effectiveness, efficiency, and sustainability. The midline report and endline quantitative report also used these criteria to frame their conclusions.

The evaluation's interpretation of the DAC criteria is as follows:

- **Relevance** focuses on the appropriateness of the design for context and purpose, in particular reviewing the validity of the TOC, particularly its key assumptions *ex ante* and whether they were based on good evidence.
- **Impact** focuses on the programme's high-level impact objective and measurement of the change which can be attributed to the programme.
- **Effectiveness** looks at whether the programme's intended outcomes were achieved, and traces through the results chains from the TOC to look at inputs, outputs, and assumptions to assess why outcomes may or may not have been achieved. Where possible, progress is reported against targets, as set out in the most recent programme logframe (2019, given in Annex C).
- **Efficiency** includes questions around cost-effectiveness, performance against budget, and timetable, governance, and performance of programme management, including use of evidence, the M&E system, and financial tracking.
- **Sustainability** relates to the likelihood that results will continue after EQUIP-T ends, in terms of beneficiaries continuing to use practices they have learned, but also activities and interventions being continued by the government. Factors deemed necessary for the results to be sustained are that the activities are considered effective and affordable, and there is stakeholder commitment and implementation capacity.<sup>18</sup>

Broadly, the questions are covered as follows:

- Relevance, impact, and efficiency are answered at the programme level, as are questions on effectiveness relating to explaining the impact.
- At the component (or sub-component) level, the questions particularly focus on effectiveness – the extent to which the outcome-level objectives were met, and the reasons why, tracing through the results chain logic from input, through output to outcome, and other reasons which may have affected the results. Questions on sustainability are included, and some components have more evidence than others in this area, including questions on affordability for the teacher in-service training and community components.

The full set of evaluation questions and the sources used to answer them are set out in Annex B.

---

<sup>18</sup> The MA notes that a further aspect of sustainability would be to consider how the interventions could be further improved, if they are adopted by the government.

### 3.2.5 Limitations

The limitations of the individual background technical studies are documented in those reports (see, in particular, OPM 2019c Section 3.5, 2020a Annex A, and 2020b Chapter 2). However, this final endline report has the following broad limitations:

- The IE was not designed with equal emphasis on all components and all DAC criteria, so there is much stronger evidence on some (such as impact on pupil learning) than others. However, this final endline report documents the evaluation's evidence on all the components and criteria for comprehensiveness.
- The pupil learning endline data was collected in 2018 (in order to avoid contamination of control schools, as discussed above), which means that it does not represent the very end of the programme's life. The programme continued for more than another year, and so it might be expected that the impact continued, or even increased, in that time. Furthermore, the specific impact of children who attended the SRP would not have fed through to Standard 3 pupil learning outcomes until 2019, and so is not captured in this data.
- Any impact of the construction sub-component at the point of data collection in 2018 is captured in the impact estimates. However, given the time taken to complete construction and open schools, this impact would likely have a longer lag time. This limitation is common in IEs, where impact may take longer to be observed than in the programme period.
- The IE uses the programme's pupil learning outcome targets, as set out in the 2019 logframe, as the basis for judging overall performance. For judging the performance of the different components, the logframe targets were regularly changed and thus do not give meaningful measures of progress. Thus, for assessing the performance of the components, the evaluation looks for evidence that the activities have been implemented on the scale intended and whether the direction of expected outputs and outcomes was found in the trend analysis. Where these trends are supported by qualitative evidence that the programme contributed to change, the assessment is even firmer. The logframe results are reported, as taken from the 2019 logframe shared by the EQUIP-T MA, and included in Annex C.
- The findings from the quantitative survey are representative of 17 districts in the original five regions, and not of all nine regions supported by the programme by the time of the endline study. Other than the pupil learning impact results, all other quantitative findings reflect trends in these districts only and are not compared with results in control districts. These results are not representative of all nine regions, for two main reasons: (1) the quantitative sample is drawn from the original five regions, which do not have the same population characteristics as the additional four regions; and (2) the implementation of the interventions was not the same in all regions – in particular, the two final extension regions had a much shorter intervention period, and not all the same activities.
- The qualitative endline findings reflect a small sample of cases selected to be best practice in focal components from the original five regions. Thus, they do not represent the experience of all schools and districts in the programme. This is both because they are purposively 'best performers' rather than average, but also because the implementation experience was different in the two extension regions.
- In the cost study, it was not possible to distinguish spending in each separate region. Thus, much of the analysis relates to the full period of implementation and gradual increase from five up to nine regions, rather than matching the scope of the quantitative and qualitative samples. Additionally, a number of assumptions were required due to limitations in how the spending data was aggregated and gaps in the monitoring data, which are explained fully in Sections 2.2 to 2.3 of the cost report.

### 3.2.6 How to read this report

Each sub-section in the findings chapters below draws on evidence from the endline technical studies and the 2019 logframe. The end of each sub-section specifies the source and where more detailed information is available. For ease of reading, the references refer to an abbreviation of the report title (EL Quant, EL Qual, or EL Cost) rather than the author and date. The report acknowledges throughout that study respondents and stakeholders have a range of differing opinions.

As discussed in Box 1 above, the quantitative trend analysis is for key indicators in programme schools in 17 original districts, and the impact estimates measure the added impact of all EQUIP-T-related interventions over and above the potential effect of the non-national LANES and Tusome Pamoja initiatives. For more information, see Section 2.5 of EL Quant. Unless otherwise specified, quantitative figures come from EL Quant. The quantitative data is therefore not representative of all nine regions, due to their different starting characteristics, and the different experience of programme implementation.

The qualitative results from endline relate to the sample of schools and districts selected as successful cases, as set out in Section 3.6 of EL Qual. Again, this is not representative of all schools and districts in the programme, because these were selected for their specific characteristics, and not all districts received the same interventions.

## 4 Programme-level findings

This chapter focuses on findings at the programme level. It starts with the question of the relevance of the design, and moves on to programme impact on learning outcomes, which is the impact-level objective of EQUIP-T. Learning outcomes are discussed further under questions of effectiveness, including the evaluation's explanation for the changes in learning outcomes, which relate to the high-level results chain given in Figure 4 in Chapter 2. Questions of programme efficiency relate to the cost-effectiveness and efficiency of spending, implementation in line with budget, and the efficiency of the management arrangements. Finally, the chapter summarises the findings.

The full evaluation framework is included in Annex B, with programme-level questions in Table 9.

### 4.1 To what extent was the programme appropriately designed to achieve its results?

**The programme itself used a TOC which was not well developed.** The EQUIP-T MA worked with a very high-level TOC, initially presented in their inception report, and later updated following the programme extension in 2017. This TOC does not provide enough detail in itself to clarify the expected pathways in the causal relationships, or the underlying assumptions that would need to hold true for the pathways to work. Thus, the TOC was clearly not well-formulated, which makes it impossible to assess how well-thought out the underlying design assumptions were.

Recognising this, OPM was asked by DFID to expand and assess the TOC at baseline. The first step in this was to start with the MA's high-level TOC and then explicitly map out the key causal pathways through which the programme expected to see change. The next step was to assess the strength of the key assumptions underpinning each link, based on the wider literature, to see if there was evidence that the particular assumption about anticipated change held in principle.

The baseline study found that most of the links between the planned programme interventions, the four component outcomes, and the ultimate impact objective in the expanded TOC were supported by existing evidence. However, the study indicated a few key areas where either the evidence was weak or other factors might undermine the effectiveness of the programme:

- There would need to be sufficient time in the school day for in-service training sessions, and these need to be linked to the performance management frameworks for teachers.
- The assumption that teachers would attend training sessions despite receiving no allowances was very fragile; other motivating factors would be needed to counterbalance the failure to provide allowances.
- The success of the teacher morale and motivation intervention would depend heavily on intervention design and context. The baseline study questioned whether the teacher morale toolkit and the implementation of a performance management system would be able to address many of the systemic inhibiting factors, including a lack of teacher housing and low salaries. Given the clear need for behaviour change among the majority of teachers, this intervention would be critical to programme impact on pupil learning, and seemed at risk. Ultimately, the programme dropped both of these interventions. The high levels of teacher and head teacher turnover was identified as a risk to the effectiveness of the various training interventions.
- The baseline found that a large proportion of children did not speak Kiswahili as their home language. The scale of this issue was not known by the MA at the time of design and is not prominent in the inception report. In this context, language acquisition would be of critical importance for the school-readiness intervention. However, whilst there was strong evidence from

the wider literature that pre-school classes can increase school-readiness on a wide range of important measures, these do not include sustainable gains in language acquisition.

- The assumption that training for district officials would lead to better planning depends on their being willing and able to change behaviours and put their learning into practice, which is highly contextual and relies on the individual attitudes, behaviour, and knowledge.
- The beneficial effects of community sensitisation can have a long lag time: the effects may not be seen within the programme life cycle.
- The broader international literature highlighted the risk of a lack of sustainability of IGAs after the initial investment, and also that there was a potential for disruption and conflict due to shifting power dynamics linked to resources.

The MA's high-level TOC includes the phrase 'especially for girls' in the outcome and impact statements, implying a particular focus on this population group in the programme design. Beyond a high-level strategy on mainstreaming gender and other equity objectives across the components, the details of how the gender goal was to be achieved are not well articulated in the design documents. A handful of the initial interventions, and more of the later interventions, had explicit gender objectives, but the programme documents lack a clear articulation and critical analysis of how these fit together to form a coherent strategy for reaching the programme's gender goals.<sup>19</sup>

Two of the initial interventions with an explicit gender focus were gender sensitisation in the teacher in-service training, and pre-service teacher training scholarships for Form 4 leavers (particularly girls). The evaluation's baseline study identified some considerable risks in the design of the latter. In particular, the assumptions required for the scholarship programme to bring more female teachers into rural schools were numerous, and at best likely hold good only in certain contexts. It also seemed likely that any change stemming from this intervention would occur only after the end of the programme. The programme dropped this intervention in the first two years of implementation, partly because of changes to the government's pre-service training programme.<sup>20</sup> The exploration of an alternative intervention – helping community teaching assistants to support rural primary schools – is mentioned in early programme annual reports but this does not have an explicit gender focus.

## 4.2 What has been the impact of the programme on pupil learning outcomes?

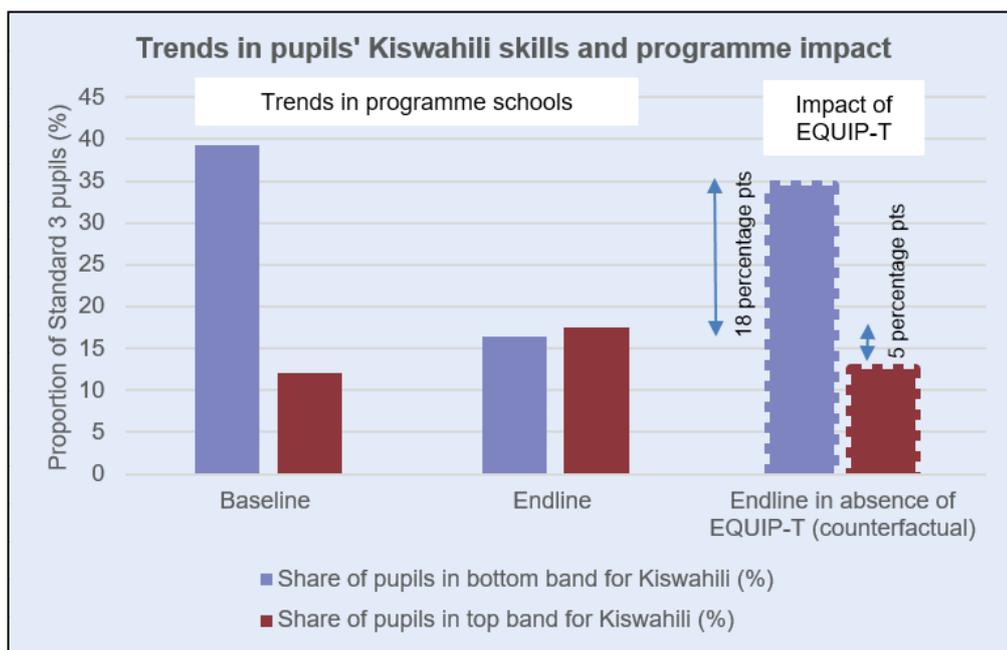
**The programme has had a positive impact on both literacy and numeracy skills for pupils in treatment schools, compared with pupils in the control schools.** Literacy and numeracy were assessed for pupils in Standard 3.<sup>21</sup> These results, representative of 17 districts from the original five programme regions, remain strong and significant across an array of estimation models and robustness checks. Figure 7 and Figure 8 show the impact results for Kiswahili and maths.

<sup>19</sup> This evaluation has reviewed EQUIP-T's initial gender and equity strategy (annexed to the MA's Inception Report in 2014), and the revised strategy (annexed to EQUIP-T's Annual Reports 2015, 2016, and 2017). The MA has advised that there was a further update to the gender strategy (a strategy for girls' education) that was approved by DFID in 2017, but this updated strategy was not part of the annual review documents in 2017 or 2018.

<sup>20</sup> The EQUIP-T Annual Report 2015 states that the reasons for dropping this were the change in the duration of the pre-service programme from two to three years, and the lack of potential impact within the programme's duration. The MA has subsequently explained that a more stringent recruitment criteria for pre-service training was another factor.

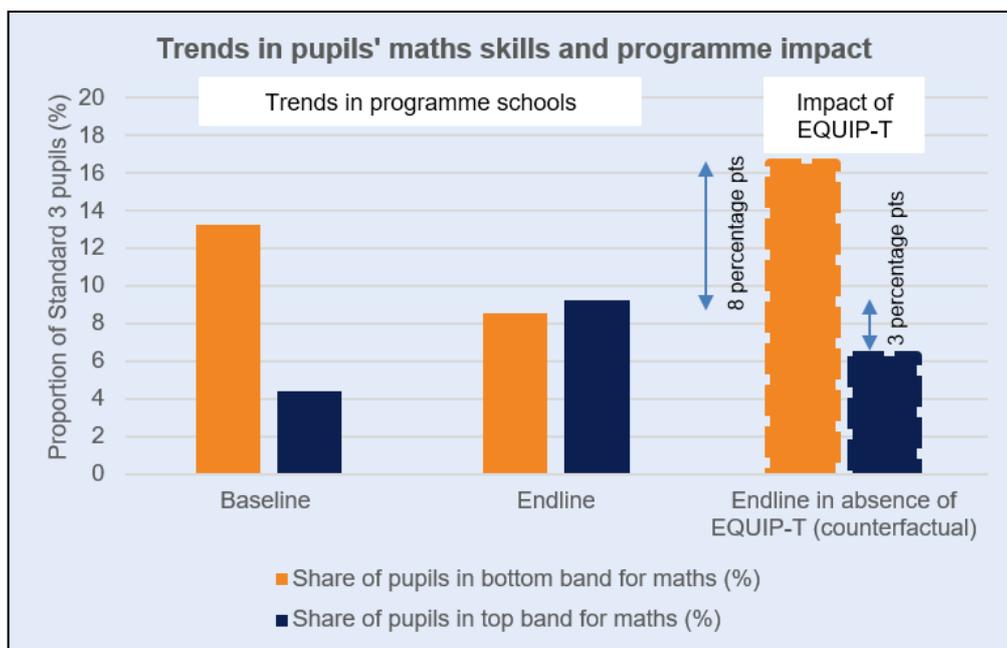
<sup>21</sup> Standard 3 pupils would have finished Standards 1 and 2, which were the main focus of the teacher quality components.

**Figure 7. Trends in pupils' Kiswahili skills and programme impact**



The positive impact on Kiswahili skills is evident in the fact that five percentage points more Standard 3 pupils are in the top achievement band, and 18 percentage points fewer in the bottom band, than would have been the case without EQUIP-T. The bands are linked to the curriculum, and the top band requires Standard 2 level curriculum skills (including reading fluency and comprehension of appropriate texts), whilst the bottom band is below Standard 1-level skills. Between baseline and midline, the programme brought pupils out of the bottom band for Kiswahili, and between midline and endline, the programme moved more children into the top band.

**Figure 8. Trends in pupils' maths skills and programme impact**



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 as it did in the control schools. With the programme, three percentage points more pupils reached the top band (Standard 2-level

proficiency, which includes four mathematical operations) than would have been the case otherwise, and eight percentage points fewer children are in the bottom band (below Standard 1 level).

It is common practice in the education literature to judge programme effect sizes using standard deviation units. There are some well-known limitations to this approach, but by this measure the programme had a 0.5 standard deviation impact on average Kiswahili scores and a 0.3 standard deviation impact on average maths scores over four years. These would typically be classified as at least moderate effects.

It should be noted that the counterfactual 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.<sup>22</sup>

Put in terms of the absolute number of pupils achieving 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 Standard 2-level Kiswahili in the 17 districts represented by the evaluation survey sample reveal an increase from 16,510 pupils at baseline to 35,596 pupils at endline – more than a two-fold increase over the four years. In maths, estimates of the number of pupils in government primary schools achieving Standard 2-level maths increased from 6,046 pupils at baseline to 18,753 pupils at endline – more than a three-fold gain.

For more information see EL Quant Vol. 1, Section 3.3.

### 4.3 Were targets for improved learning outcomes achieved? Did learning gaps narrow in favour of marginalised groups (girls, non-Kiswahili speakers)?

**Despite the positive impact, the programme did not meet its logframe target for learning outcomes in literacy, and narrowly fell short of the target for numeracy. However, these targets relate to pupils at the upper end of the performance distribution; there were further improvements for the lower-performing pupils, and a narrowing of gaps between groups.**

The impact indicator in the programme's logframe is the percentage of pupils meeting Standard 2-level literacy and numeracy by Standard 3. The targets for 2019 are 25% in literacy and 10% in numeracy, for boys and girls. These are up from a baseline of 12% and 4%, respectively. The endline data is for 2018 rather than 2019, and at this point neither target was met: the proportion of pupils achieving Standard 2-level literacy was 18%, and in numeracy it was 9%. Girls perform better than boys in literacy, and the reverse is true in numeracy, as was the case at baseline.

Despite not meeting the target, learning outcomes have substantially improved in treatment schools. The proportion of pupils meeting Standard 2-level literacy increased from 12% at baseline to 18% at endline. Those in the bottom band, in other words below Standard 1 level, fell from 39% to 16%. For maths, the share meeting Standard 2 level increased from 4% at baseline to 9% at endline; those below Standard 1 fell from 13% to 9%.

The progress in terms of absolute numbers of children meeting curriculum expectations is even more pronounced, due to the steep increase in enrolment intake between baseline and midline, leading to

<sup>22</sup> Section 2.5.1 in the EL Quant Vol. I explains the key assumptions that the evaluation has made in interpreting the impact estimates.

far higher enrolment in Standard 3 by endline. A higher share out of a higher total cohort of pupils means that more pupils, by a factor of two or three, are meeting Standard 2 levels at endline than at baseline in the sampled districts, as discussed in Section 4.2 above.

Although this progress is very encouraging, it is important not to lose sight of the extent to which pupils are still behind curriculum expectations. This evaluation finds that 82% of pupils are still below the expected level in literacy, and 91% of pupils below that for numeracy. This scale of learning deficit is in line with results reported for the government's national Education Program for Results which are based on similar methods and performance criteria.<sup>23</sup> The National Examinations Council of Tanzania (NECTA), in a recent early grade learning study, used different performance criteria and methods, and reported far more positive results.<sup>24</sup>

The evaluation also measured changes in the relative levels of learning outcomes for marginalised groups within the EQUIP-T schools. This is trend analysis and does not indicate a change over and above what would have happened without EQUIP-T:

- 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 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.
- Pupils from richer backgrounds have continued to outperform pupils from poorer backgrounds since baseline, and whilst the gap has narrowed modestly for Kiswahili, it has grown for maths.

For more information, see EL Quant Vol. 1, Sections 3.3, 3.4, 3.5, and 3.6, and Annex C (Logframe Sep 2019).

## 4.4 What factors explain these results?

### 4.4.1 What explains the changes in learning outcomes?

**This evaluation finds that part of the improvement in learning outcomes is attributable to the impact of EQUIP-T over and above a counterfactual, and part can be explained by national trends.**

**In terms of EQUIP-T's additional impact, the teacher in-service training is likely to have made the biggest positive contribution to impact on pupil learning.** The positive impact on pupil learning follows the interventions on early grade teacher training closely: 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.

<sup>23</sup> The Progress in Results Framework for the Education Program for Results is reported annually. This shows the percentage of Standard 2 pupils that reach national targets for reading fluency in Kiswahili and for minimum numeracy skills. See MoEST and PO-RALG (2018, p. 153).

<sup>24</sup> NECTA carried out a sample-based assessment of 3Rs skills among pupils who had completed Standard 2 in 2017. The study reported the percentage of pupils with good performance in each skill area. See NECTA (2018).

Implementation of the teacher in-service training has occurred broadly as intended, and participation of teachers is relatively high.

Teachers, head teachers, and WEOs describe the techniques gained from in-service training as leading to improvements in learning outcomes, especially in the 3Rs. Teachers credit the training as helping them identify pupils who are struggling with literacy and numeracy, and that the mix of techniques learned from the training means that more pupils can be reached to help them fully understand the topics. There was a notable deficit in teachers' skills before the programme; EQUIP-T addressed this need directly.

Meanwhile, the other components have a less direct link to improving pupil learning, which makes their impact on the improvement in learning less clear. Whilst they contribute to a supporting environment for effective teaching and learning, they have been less successful in terms of achieving the intended intermediate outcomes as set out in the programme TOC (see the results chain in Figure 4). This raises the question of whether the potential impact on pupil learning could have been substantially greater if other components had been more successful.

**Some of the improvement in learning outcomes is explained by common trends experienced across the country.** Findings from the nationally representative 3Rs (early grade reading and maths assessments – EGRA/EGMA) study at the time of the midline showed 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. The results for the EQUIP-T control sample were similar. The trend at this time was likely due 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 for maths. The other national factor which emerged from the midline research was *hapa kazi tu* (translated as 'just work') – the slogan introduced by the new government, which was credited by respondents across the midline qualitative work with influencing teachers' work ethic, and also with increasing the level of oversight and monitoring of schools and teachers.

Meanwhile, external influences are likely to have put negative pressures on the programme. The introduction of fee-free basic education led to a surge in enrolment in 2016, which has been sustained, and the government's professionalisation policy (see Section 2.4.2) caused high turnover of head teachers and WEOs in 2017. Indeed, a more recent national 3Rs assessment held in 2017 shows scores in Kiswahili and maths falling back to 2014 levels. The evidence suggests that EQUIP-T has helped to prevent a similar learning decline in its programme schools. These negative pressures could also be reasons why the programme failed to achieve its endline learning outcome targets, at least by 2018 (as discussed in Section 4.3).

**There is some evidence that EQUIP-T has contributed to the larger gains for girls.** Teacher interactions with pupils have become more gender-balanced. This improvement was seen between baseline and midline, when the first four teacher modules introduced gender-responsive pedagogy (and a specific gender-responsive pedagogy training session was rolled out after midline). In the qualitative endline interviews, teachers explained that the training led them to be more gender conscious in classroom seating arrangements and their interactions with pupils.

Similarly, teachers interviewed for the qualitative endline reported that the in-service training helped them learn how to support pupils who do not speak Kiswahili as their mother tongue, which may have helped slightly lessen the gap for these children. Teachers report that they still find pupil language to be a major challenge, however.

For more information see EL Quant Vol. 1, Sections 3.3 and 3.4, and EL Qual, Section 5.1.

#### 4.4.2 Did the demand-side assumptions required for the programme to be effective hold?

**The positive results for learning outcomes were achieved despite some worsening in pupils' backgrounds, which suggest the programme could have potentially had even higher impact if these background factors had not changed.** Pupil backgrounds and home support factors can inhibit (or enhance) the link in the TOC between improving schooling and achieving the desired learning gains.

Some changes in pupils' backgrounds likely occurred due to the fee-free policy announcement in 2016, which led to an influx of new pupils. Pupils' economic circumstances have worsened: close to 40% of Standard 3 pupils come from households that fall below the national poverty line, up from 33% at baseline. This has various ramifications, as it means more pupils have economic responsibilities outside of school, and fewer pupils have eaten before coming to school, affecting their readiness to learn. The gender balance and average age of Standard 3 pupils has not changed significantly, although the share of over-age pupils has increased, likely due to over-age pupils entering school for the first time three years earlier, shortly after the fee-free policy announcement.

In terms of educational support, more pupils have support for learning at home than at baseline, either through reading materials at home or the availability of someone to help pupils with homework. Teachers still feel that parents lack awareness of the importance of education and this disadvantages their children. The proportion of pupils who attended pre-school has not changed since midline (it was not recorded at baseline), at around three quarters. The cohorts who have attended the EQUIP-T SRP would not have reached Standard 3 until the year after the endline survey, and thus any additional impact from this intervention was not picked up in this evaluation's results.

Pupil absenteeism has reduced, but is still high. For pupils in Standards 1 to 3, daily absenteeism fell significantly, from 34% to 28%, since baseline, and there has been a slightly larger decline for girls than boys. Teachers and parents report that (lack of) pupil attendance and engagement is an issue for learning, and is affected by distance to school, chores or work, and hunger.

For more information see EL Quant Vol. 1, Section 3.2 and EL Qual, Section 5.2.

### 4.5 To what extent was the programme cost-effective?

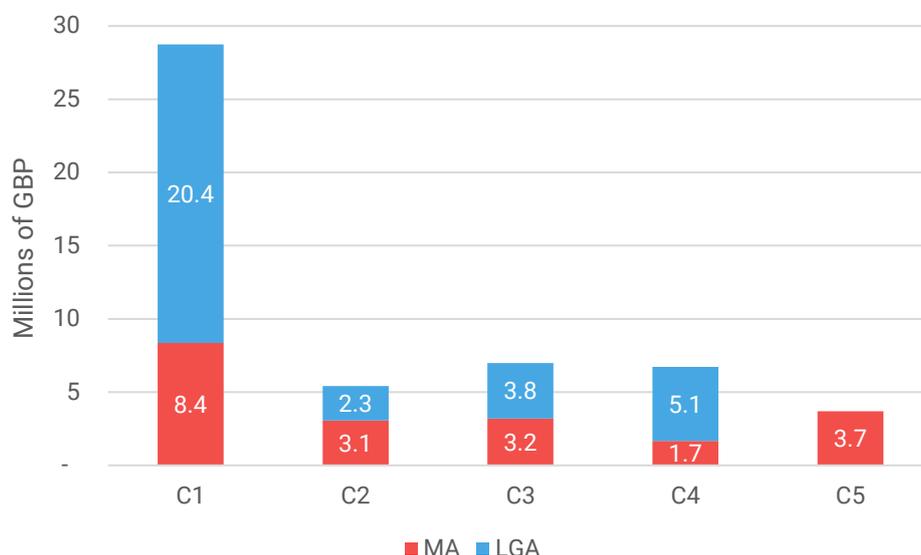
#### 4.5.1 What did the programme cost to deliver, and what are the main cost elements?

**The programme as a whole cost over £72 million to deliver between its start in February 2014 and June 2019, meaning 90% of the total budget had been spent.** More than one quarter of total spending was on technical assistance, leaving three quarters (£52 million) as PSA spending at MA and LGA level. Of this, over 60% (£32 million) was spent through LGAs and the rest by the MA. Out of a budget for decentralised LGA funds of £37.2 million, it is problematic that 15% was still left to spend when there were only six months of the programme remaining. From that point, the MA was reducing its support and supervision: it creates a risk that the remaining spend offers poor value for money (see KPMG, 2019), and if not spent at all then misses an opportunity for impact.

The breakdown of PSA spending by component is shown in Figure 9. Component 1, targeted at improving access to high-quality education, made up 56% of PSA spending (£29 million), and the largest portion of this was the teacher professional development model, using district-level, ward/cluster and school-based in-service training. School construction, which was introduced after the programme extension in 2017, also contributed a substantial portion of overall costs.

Components 3 and 4 (district planning and communities) were both the next largest, having cost 14% (£7.0 million) and 13% (£6.7 million), respectively, of the total PSA spend. Component 2, SLM, was 10% of PSA spending. The M&E component cost a total of £3.7 million and included the purchase of tablets for schools, as well as annual monitoring surveys.

**Figure 9. Total PSA expenditure, by category and component**



Source: EQUIP-T data. Includes spending from February 2014 to June 2019 inclusive.

For more information, see the EL Cost, Section 3.

#### 4.5.2 How cost-efficient has the programme been in delivering (selected) outputs? What can be said about the cost-effectiveness of the programme?

**Cost-efficiency analysis shows that there have been large variations in implementation between LGAs, indicating challenges in capacity, support, or financial data reporting and monitoring.** Cost-efficiency analysis was conducted by calculating the average unit costs of activities for the entire programme duration to date.<sup>25</sup> It focuses only on spending by the programme, not any additional costs incurred out of other budgets or out-of-pocket costs. On average, the programme has spent £5,900 per school on support to teachers and the SRP. The programme has cost around £1,400 per school for SLM activities, to train head and assistant head teachers, and WEOs. Support to strengthen district planning and management has cost £156,000 per LGA. Within this, LGAs spent on average nearly £2,500 per WEO on WEO grants. The community component cost a total of £1,800 per school, which included training for SCs, PTP grants and IGA grants, and school noticeboards.

Cost-efficiency analysis of LGA expenditure reveals some substantial variation in unit costs across regions. For example, Simiyu spent £370 on providing one teacher with the full package of in-service training, whereas Tabora spent £760. Some activities were based on per school fixed unit costs and so should be consistent across regions, such as the PTP and IGA grants, yet there was variation. For example, the spending on IGA per recipient school was £420 in Mara and yet over £930 in Lindi and Dodoma, when it should be £490 everywhere (see Figure 10 and Figure 11 below). This evaluation understands that all LGAs should have implemented the same activities, using the same budgeting

<sup>25</sup> The analysis adjusts for the variation in the number of regions (and LGAs, schools, and so on) over the period.

formulae. Thus the variation in unit costs could relate to a number of underlying factors: varying levels of capacity to implement the volume of activities in LGAs; a failure to implement according to guidelines; legitimately using varied models with differing costs; withheld funds due to unresolved misspending; or errors in the LGA spending data provided by the MA.<sup>26</sup> More routine and regular monitoring of this sort – using average unit costs to verify expenditure data and beneficiary data – is important for programme managers to understand what is going on and respond to problems if needed.

**Figure 10. LGA unit cost: average IGA grant per recipient school**



Numerator: spending on IGA grant. Denominator: the number of schools with IGA grants.

**Figure 11. LGA unit cost: average PTP grant**



Numerator: spending on PTP grant. Denominator: number of schools in the region.

Source: EQUIP-T spending and M&E data. Spending data is for LGAs from January 2016 to June 2019.

**This evaluation has not estimated cost-effectiveness**, which would compare the impact of the programme on pupil learning, as estimated by the quantitative evaluation, with the associated incremental cost. The methodological requirements for this analysis were explored following the midline study and presented at the 2017 Education and Development Forum (UKFIET) Conference. There are two major challenges:

- Cost-effectiveness ideally measures the marginal cost required to deliver one unit of impact – in this case, the impact would be measured as the average change in learning outcomes in standard deviations. There is a theoretical challenge with defining what the marginal cost is. The learning outcome is achieved by a cohort of pupils in Standard 3; however, it represents a result of the interventions and changes in the school over four years (mid-2014–mid-2018). Of these four years of interventions, which each cumulatively contribute to one cohort year of impact, which of their costs should be identified as leading to the change? In addition, a large number of interventions took place and it is not possible to robustly estimate the contribution of each one to the learning outcome result. This could mean that the costs of an activity are included – raising the total cost – yet they were not at all effective. This, of course, would lower the apparent cost-effectiveness of the whole programme.
- An equally serious challenge, related to the first, is that the programme's spending data is not adequately captured and recorded to allow the actual marginal cost to be estimated. Even at aggregate level (say, the marginal cost per region), the spending data does not separate out the initial activity development costs, or categorise by region. The evaluation team discussed this problem with the MA when the programme expanded from five to seven regions. The programme

<sup>26</sup> The evaluation team was provided with six iterations of the dataset in response to queries over anomalies pointed out by the evaluation team.

has not been implemented in the same number of regions throughout its lifetime, so dividing the total costs by numbers of beneficiaries is not an accurate reflection of what it cost to deliver the programme to each beneficiary.

Whilst a cost-effectiveness ratio has not been calculated, the costs and effects can be looked at in broader terms. The programme cost over £72 million to deliver, and in 2019 reached 5,192 primary schools with close to 3.2 million pupils enrolled. The total number of beneficiaries over the entire programme is not provided in the MA's monitoring data, which is a limitation. If one seventh of the school population was replaced each year, this would give a total number of pupil beneficiaries of around 5 million (though this is a very crude estimate). Again (also a crude estimate), a total spending of £72 million over five years implies £14.4 million per year, which over 3.2 million pupils would be £4.50 per pupil per year. The number of pupils in the early grades, who were the ultimate beneficiaries of the teacher in-service training activities, was not tracked by the MA. If only those pupils were included, it would give a higher annual spend per pupil.

For more information, see the EL Cost, Section 3.

## 4.6 To what extent was the programme implemented in line with its agreed budget and timetable?

### 4.6.1 To what extent were adequate and complete budgets prepared and used to guide expenditure?

**A comprehensive budget for EQUIP-T broken down by meaningful categories and time periods is not readily available, making it difficult to assess the budget planning process and execution. This lack of information in itself is an inefficiency and limits transparency and accountability.**

The efficiency of budget execution against the planned budget is one aspect of the concept of operational efficiency; this aspect looks at whether the programme spent as intended, whether spending was in line with budgets, where there were over- or under-spends, and the reasons why. For adaptive programmes, such as EQUIP-T, which are intended to be responsive to context, the reasons for low budget execution rates or for large within-year budget revisions may well be valid, but these need to be well documented so that they can be assessed. If budgets are not prepared and agreed from the outset, there is no framework to ensure that expenditure is guided by the objectives of the programme, or to assess whether expenditure and implementation are on track.

Over the course of implementation, the process of preparing budgets changed. For the first three years of the programme, the MA had annual PSA budgets, and similarly the decentralised LGA funds had annual budgets. However, from the 2017 programme extension onwards, although there was a detailed budget for the whole extension period (2017–2020) by activity, there were no agreed budgets set for a medium-term period (e.g. annual or tranche) at the level of the 10 sub-components or by LGA and Epicor code.<sup>27</sup> Budgets were not detailed using the same type of categories (codes) as the expenditure data, and were not held as a cumulative running total budget. Furthermore, there is no budget for the entire programme lifetime for PSA funds broken down by component. Since the extension, the MA has produced a rolling monthly workplan and forecast for MA spending as an internal tool, but has not been required to set and report on medium-term budgets to DFID. LGA

<sup>27</sup> Epicor is the government's integrated financial management information system at sub-national level.

spending has been budgeted for individual tranches, but no cumulative tracking of these budgets (over seven tranches) exists categorised by LGA and Epicor code.

The lack of reported budgets since 2017 is a weakness in the programme's management: it is not possible to track transparently whether the programme was achieving its spending objectives annually and planning efficiently, or why changes in spending decisions were made. Budget execution cannot be meaningfully assessed against a plan which changes monthly.

For more information, see the EL Cost, Section 3.

#### 4.6.2 To what extent were budgets fully executed, and how did the rate of budget execution vary between categories/activities?

**The MA spent 97% of its PSA budget with less than 10% of programme duration still to complete. In terms of funds decentralised to LGAs, the range in average levels of spending indicate that budget execution has been incredibly varied.**

Up to June 2019, execution against the MA's PSA budgets was high, as would be expected near the end of the programme, at over 95% for all components (except component 4, which was 85%). This high level of budget execution is expected at the end of the programme, given that the final months were intended to be for closing down the project, and the programme revised its budgets regularly. The low execution on component 4 may be related to work on supporting the roll-out of girls and inclusion activities, which were delayed given the delay in disbursing tranche 7 to LGAs. As a result, the MA would have delayed its own activities to support this roll-out. A review of budget execution in the first three years of the programme reveals that the programme started in 2014 by hugely over-estimating its budgets and under-spending; by 2016, the programme was actually spending more than planned in its annual budget.

For the LGA decentralised funds, the lack of cumulative or total budgets per LGA and by Epicor code makes it impossible to estimate final budget execution rates. One way of looking at relative spending is the average spending per school for the LGAs. Looking at the seven regions which all received funds for the same period, average spending varies dramatically, from £7,400 per school in Lindi to only £5,700 in Simiyu. This evaluation is aware of no programme design to explain such a large variation in spending. The lower average spend in Singida, at £3,300, is explained by the region joining the programme more recently.

Budget execution for 2018/19, which did have an annual budget,<sup>28</sup> shows substantial variation across regions and components in execution rates. For regions, this varied from as low as 52% in Simiyu to as high as 117% in Dodoma, and for components, from 19% on community and school partnerships up to 365% on 3Rs in-service training.

Inconsistent budget execution at the level of LGAs relates to a number of reasons: funds carried over from the previous year; withheld disbursements due to previous spending issues; delays in transfers – an issue at the level of central government – and implementation capacity problems at the LGA level; and possibly changes from initial plans set by the MA. Taken together, the planned budgets appear to have been overly optimistic and demonstrate the challenge of budgeting in this operating environment.

For more information, see the EL Cost, Section 3.

<sup>28</sup> The reason why the financial year 2018/19 was subject to an annual budget but other years were not is unclear.

## 4.7 How well did key management arrangements support programme implementation?

### 4.7.1 Was evidence and a TOC used to guide the programme's design and adaptation?

The EQUIP-T MA TOC is more of a high-level summary of the programme's content and overall ambition than a detailed formulation that could be used for critical reflection and programme adaptation. This applies to both the MA's TOC developed at inception (2014) and the later version developed prior to the programme extension (2017).

EQUIP-T's design changed over the first few years of the programme, with some activities being removed, whilst some of the initial activities were deepened – for example, the teacher in-service training programme evolved to include more elements, whilst scholarships for rural candidates to attend pre-service training were dropped. Apart from adjusting the content of the programme, the move to decentralise the main funding of EQUIP-T activities to LGAs was a significant departure from the initial design.<sup>29</sup> The programme extension in 2017 brought some large new sub-components, including one focused specifically on the needs of girls and other marginalised children in accessing and participating fully in schooling. This marked a departure from EQUIP-T's initial strategy of mainstreaming gender and equity interventions across the programme.

Based on the types of statement included in programme documents to justify programme design changes, there is no evidence that these changes are the result of a TOC-driven reflection process. Apart from the introduction of the construction element, and possibly the change in the programme's gender and equity approach,<sup>30</sup> the main design changes largely appear to have been driven by the need to solve challenges that occurred during implementation related to government institutional constraints. Whilst this is a sign of the MA adapting the programme based on learning more about what works and does not in the context, the potential consequences of these changes for the achievement of programme outcomes and impact is not systematically documented or taken into account in any adaptation of the wider programme design.<sup>31</sup>

That said, the baseline and midline IE reports, which were produced before the final programme design covering the extension, used a TOC developed for the purpose of evaluation to assess the programme.<sup>32</sup> Some of the resulting evaluation recommendations have been taken up by the MA in

---

<sup>29</sup> From 2015/16, funds were transferred to LGAs for them to manage, with technical support from EQUIP-T, rather than the EQUIP-T MA spending funds directly. Under the original design, this would have happened more slowly, with greater support from EQUIP-T to improve LGA PFM skills prior to funds being decentralised. This change was introduced in recognition of the fact that such a large volume of funds and interventions could not be implemented centrally, and working through decentralised government systems was expected to increase local ownership and sustainability.

<sup>30</sup> The rationale for making a separate sub-component focused on girls and other marginalised children, rather than the initial mainstreaming approach, is not explained in the programme's extension document beyond asserting that 'there is a clear need for a greater focus on some of the biggest barriers to girls maximising their potential' (EQUIP-T MA, 2017, p.20).

<sup>31</sup> For example, the consequences of no longer requiring LGAs to reach specified PFM performance benchmarks before being given authority to manage EQUIP-T funds most likely raised the risk of not achieving the original LGA PFM capacity building targets (at least within the original timeframe) and thus its expected contribution to programme outcomes.

<sup>32</sup> As set out in Chapter 2, the evaluation team developed a more detailed TOC for the purposes of the evaluation based on their understanding of how the programme was intended to work. The EQUIP-T MA TOC was used as a starting point, and

adjusting the programme design. The issue of growing Standard 1 class sizes was highlighted in the midline evaluation report, for example, and it is clear that the additional construction component was designed to lessen this problem. In other cases, it is less clear whether key evidence from the evaluation has fed sufficiently into programme adaptations. In consecutive rounds of this evaluation, the results have shown that the overwhelming majority of pupils in the EQUIP-T districts do not speak Kiswahili as their main language at home and that these children have particularly low levels of achievement. At midline, the evaluation noted that 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 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 by home language (Kiswahili or other), it is plausible that a more specific intervention could have helped reduce the gap further.<sup>33</sup>

Some other major threats to programme success, such as high levels of teacher absence from classrooms, and high levels of turnover among teachers, head teachers, and other education officials, were consistently flagged in all the evaluation reports, yet these issues were rarely mentioned in EQUIP-T Annual Reports, and no explicit interventions were put in place to try to mitigate these.<sup>34</sup>

The Management Background note contains further information on this.

#### 4.7.2 Were strong M&E systems in place to monitor and adapt implementation, and provide accountability?

EQUIP-T's framework for monitoring its progress at output level and above was its logframe. It is questionable whether the set of logframe indicators alone provided sufficiently comprehensive data for a programme of the size and complexity of EQUIP-T to enable critical reflection on progress. Moreover, the degree of changes in the logframe indicators from year to year seriously compromised their ability to act as monitoring tools beyond the very short term.<sup>35</sup> For tracking input, activity, and beneficiary data, the MA relied on manual monthly narrative reporting from its regional team leaders, and periodic requests to them for specific data to populate programme Fact Sheets. Early attempts to implement a programme management information system in a database format were unsuccessful.

The quality of the key data sources used to monitor the programme is very mixed. The main data source used to report logframe indicators at output level and above is the annual monitoring survey. This survey was conducted over the first four years of the programme by an external firm using a consistent design, fieldwork, data management, and reporting process that included many of the type of quality control measures that would be expected in a survey of this type.<sup>36</sup> By contrast, any quality

---

the version developed for the evaluation articulated key causal pathways and identified assumptions under which each element of the causal chain could be expected to hold.

<sup>33</sup> The MA has explained that the SRP was partly designed to help non-Kiswahili-speaking children to prepare for school, and to support Standard 1 teachers in helping these children transition to school. However, this is not well documented in programme reports, and there is no evaluation evidence available on this.

<sup>34</sup> The MA explained that these types of risks were tracked in the programme's risk register.

<sup>35</sup> By Year 6 of the programme, only three of the original 27 indicators were still in the logframe and had at least some populated achieved values.

<sup>36</sup> The annual monitoring survey was replaced by a survey carried out by WEOs for the final two years of the programme, as part of EQUIP-T's institutional strengthening strategy. The MA explained that this change was also made because of its concerns about the quality of the annual monitoring survey.

control processes applied to the data included in the periodic programme Fact Sheets (beneficiary, input, and activity data) are not documented in the sheets themselves or in other documents shared with the evaluation team. The data gaps and multiple values found in some of the cells in these sheets are further signs of problems with the quality of this source.

EQUIP-T's Annual Reports provide the main documentation for both accountability and learning purposes. The first two reports were fairly comprehensive and transparent, and provided detailed performance monitoring data and narrative explanations. From 2016 onwards, the content of the reports narrowed considerably, and their formats changed each year, giving far less detail on programme progress beyond the immediate year of reporting.<sup>37</sup> The scale of changes in the logframe indicators over this period also meant that there was very little reporting of performance trends possible. Over the last four years of the programme, it has become increasingly difficult to trace progress against longer-term work plans and expected outputs, considerably reducing transparency and the ability to use the reported information to hold the programme to account.

The Management Background note contains further information on this.

#### **4.7.3 Were strong financial systems in place to monitor spending and provide accountability?**

The structure of the chart of accounts for EQUIP-T seriously limits the type of spending analysis that can be carried out, including types of value-for-money analysis and an estimation of the cost of scaling-up the programme. The availability of EQUIP-T MA budgets as lifetime totals only, apart from during the first three years of implementation, when annual budgets were kept, makes it impossible to assess budget planning and execution in a meaningful way. Similarly, the absence of LGA PSA budgets categorised by LGA and the government chart of accounts limits the assessment of budget planning and execution for decentralised funds.

The protracted process of obtaining EQUIP-T spending data from the MA, particularly LGA spend (which accounts for just under half of the total EQUIP-T budget), and the errors in the datasets that have been provided to the evaluation team are not signs of a well-functioning and reliable financial management system.<sup>38</sup> A recent fiduciary risk assessment focusing on the decentralised funds rated the programme as a 'substantial risk', highlighted that 'financial reporting needs to be much improved' (KPMG, 2019, p. 16), and made the recommendation that the 'EQUIP-T Managing Agent should clarify arrangements for review and quality assurance of Quarterly financial reports before they are submitted to DFID' (ibid, p. 157).

Financial transparency has diminished over the programme period, at least to the wider annual review audience. Financial data (budgets and spending) included in EQUIP-T Annual Reports has become far less detailed over time, and thus these reports have become increasingly less useful as tools for the accountability of programme spending.

The Management Background note contains further information on this.

<sup>37</sup> The MA explained that they were requested to make the reports easier to read and more suitable for a wider audience.

<sup>38</sup> The evaluation team was provided with six iterations of the dataset in response to queries over anomalies pointed out by the evaluation team.

## 4.8 Programme-level summary

Table 3. Summary of findings at programme level

### To what extent was the programme appropriately designed to achieve its results?

**The design was likely to achieve results but with some notable risks.**

The original programme design was largely likely to achieve its results based on the strength of evidence that there were links between interventions, outcomes, and intended impact. However, there were a number of risks in the assumptions which could undermine effectiveness – the needed motivation to attend training and change behaviours, as well as the lag times in changing community relationships.

### What has been the impact of the programme?

**The programme has had a significant positive impact on learning outcomes for pupils in the early grades.**

Pupil learning in Kiswahili and maths has improved, shifting pupils out of the lowest-performing bands towards the expected level of curriculum attainment for Standard 2 (the highest band). The improvement was larger in Kiswahili than in maths, likely due to the longer focus on Kiswahili.

### Was the programme effective in meeting its targets for improved learning outcomes, and why?

**The programme did not meet its logframe targets, but the improvement is notable nonetheless.**

Whilst the programme did not quite meet its logframe targets (which represents one measure of the improvement in learning outcomes), the programme has made a significant contribution to preventing children from being left far behind curriculum expectations. This is particularly impressive given the increase in enrolment, meaning that far more pupils in absolute terms are now meeting curriculum standards. The largest contributor to EQUIP-T's impact on learning outcomes over and above the national trends is the teacher in-service training, which was implemented largely as planned and has the closest direct effect on pupil learning.

### To what extent was the programme efficient?

**Whilst the programme has delivered and achieved a large amount, there are some serious inefficiencies in the way it was managed.**

The programme cost over £72 million to deliver up to mid-2019, and in 2019 it reached 5,192 primary schools, with close to 3.2 million pupils enrolled. Variation in unit costs across LGAs suggests that implementation was not consistent. Given the adaptability needed to respond to the implementation and spending of LGA grants through government systems, the programme lacked detail in many areas of its budgets, which make it difficult to track implementation progress and provide accountability. Similarly, weaknesses in the coding structures and errors in data provided suggest the financial management was not functioning well.

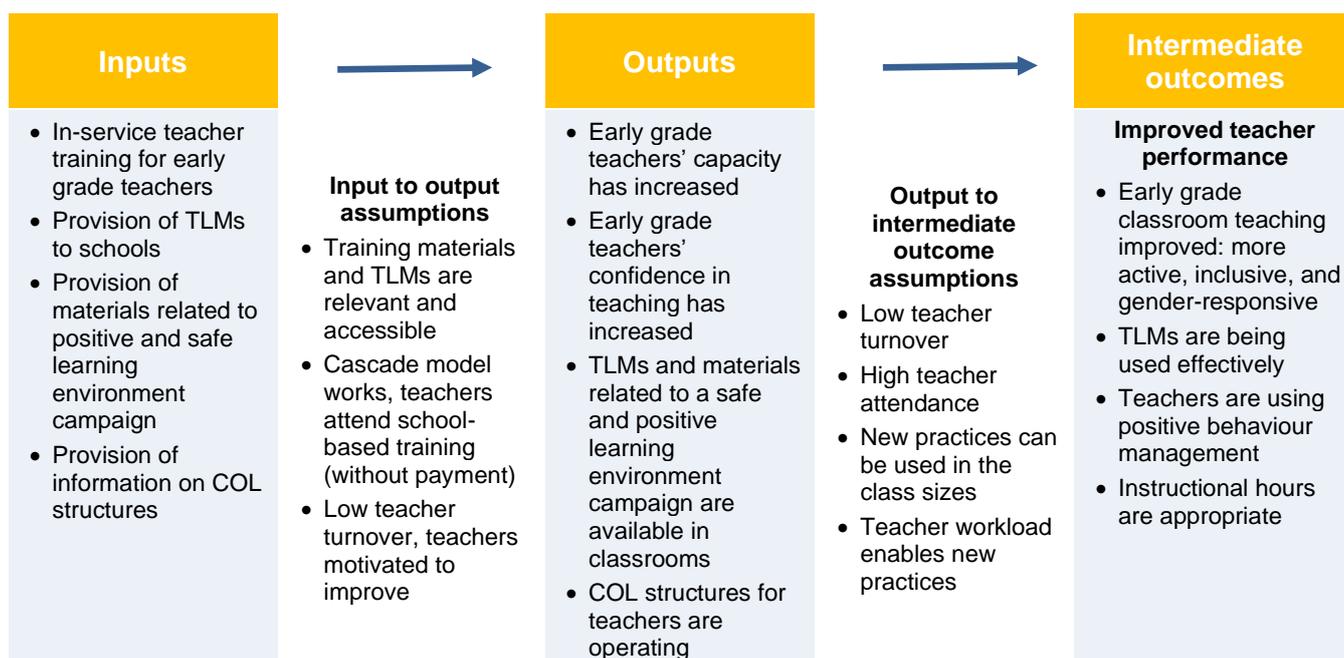
The programme made many adjustments over its duration, driven largely by the need to solve challenges that occurred during implementation related to government institutional constraints, rather than the result of a TOC-driven reflection process. Furthermore, the degree of changes in the logframe indicators from year to year seriously compromised their ability to act as monitoring tools beyond the very short term. The programme was unsuccessful in setting up a management information system database, and its periodic Fact Sheets, used for monitoring, contain a number of issues.

## 5 The teacher component

This chapter summarises the findings of the teacher component – which focuses on sub-component 1A. The questions relate to the effectiveness and sustainability of the sub-component. This chapter briefly begins with a summary of the evaluation results chain (a sub-set of the TOC) which was used for assessing effectiveness and sustainability.

Under sub-component 1A, the inputs are focused on the early grades (Standards 1 and 2) and include: in-service teacher training; the provision of TLMs; the provision of materials related to a positive and safe learning environment campaign; and the provision of information on COL<sup>39</sup> structures. If a number of assumptions hold, the activities are expected to lead to outputs which include an increase in teachers’ capacity and confidence, the availability of materials in the classroom, and operational COL structures. These, in turn, lead to the intermediate outcomes: an improvement in early grade classroom teaching; effective use of TLMs; teachers using positive behaviour management; and appropriate instructional hours (due to teachers attending their lessons more often). The key assumptions in this results chain are that the materials are relevant for the context and accessible; the cascade model is effective in passing on knowledge; teachers attend in-service training without material incentives; teacher turnover is low; and motivation and attendance are high. In terms of the relevance of the materials, teachers are assumed to be able to use the practices in their classrooms and fit them into their workloads.

**Figure 12. Main results chain for EQUIP-T sub-component 1A teacher performance**



EQUIP-T’s logframe indicators which relate to this results chain are: Outcome indicators (1) Early Primary Teacher Performance in Pedagogy: score of assessment of classroom observation of teacher methodology (sex disaggregated); (2) percentage of teachers using gender-responsive pedagogy in their classroom teaching. Output indicators: percentage of schools where teachers are engaged in improving their teaching through active in-service training groups.

The full evaluation framework is included in Annex B, with component 1 questions in Table 10.

<sup>39</sup> COLs are 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 CPD.

This evaluation does not focus on sub-component 1B (the SRP) or sub-component 1C (the construction). The first cohort of children who attended the SRP would not be in Standard 3 until 2019, and thus the improvement in average learning outcomes due to this intervention was not picked up in 2018. In terms of construction, if this had any impact on quality of education, and pupil outcomes, by 2018, this would be included in the pupil learning impact estimates. However, it is likely the construction had not yet affected outcomes in 2018. The intention of the sub-component was to complete four unfinished classrooms, and build four satellite schools (each with two classrooms), in each LGA. An average LGA is responsible for close to 100 primary schools, with close to 60,000 pupils enrolled; thus, the size of any eventual impact of the construction will be likely to reflect the scale of this intervention.

## 5.1 Did the teacher component achieve its outcome targets for teacher performance?

**The programme has two 2019 logframe targets for measuring teacher performance – effective teaching practice and gender responsive pedagogy<sup>40</sup> – neither of which have been reported on since 2016. However, the evaluation collected data against similar indicators; these show that performance against measures of teaching effectiveness and gender inclusion has been mixed.**

**The evaluation finds that a minority of teachers demonstrate a range of positive teaching practices in the classroom and that this has significantly worsened over time.** The evaluation measures 14 teaching practices that are considered to characterise positive teaching and classroom practices. At baseline, teachers were demonstrating at least seven positive teaching practices, out of the measured 14, in the majority of lessons (68%), but this declined to 56% of lessons at midline, and to 40% at endline. This finding suggests that there is considerable scope for the further strengthening of general pedagogy. The use of positive teaching practices in the introductory and concluding lesson stages has significantly and substantially worsened since both baseline and midline. Changes in the use of positive teaching practices during the middle stages of a lesson have been mixed. The decline in the use of many of the positive teaching practices is likely to be the result of a number of factors, including the considerable increase in class sizes and the introduction of the new 3Rs curriculum, which changed the structure of Kiswahili and maths lessons. Both of these are discussed in more detail in Section 5.2.4 below.

Within these overall disappointing results, there are some positive findings. 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). There have also been improvements in the share of lessons where teachers ask pupils to demonstrate in front of the class, with teachers providing written or verbal feedback to pupils on their individual work. 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%), but there remains a sizeable minority of lessons (44%) in which no materials were used at all. Furthermore, despite the majority of schools receiving the EQUIP-T supplied materials, these are not being used: only 4% of observed Kiswahili lessons used the 'big books' or 'read-aloud' books.

Other indicators of effective teaching relate to practices outside of lessons, and there are mixed findings: the proportion of lessons which have a written lesson plan has significantly worsened since

<sup>40</sup> The first logframe target relates to effective teaching practices and is not well defined. It refers to a 'score of assessment of classroom observation of teacher methodology'. The second target is the percentage of teachers using gender-responsive pedagogy in their classroom teaching. Neither have been measured and reported on since the mid-term review in 2016.

baseline and midline. On the other hand, the qualitative study showed that teachers in the sampled schools feel that they have learned how to plan and prepare for lessons in a way that aids teaching, though carrying out lesson planning is also largely dependent on teachers' workload (more lessons and large class sizes affect available time and volume of marking). There has been some improvement in the use of regular pupil assessment by teachers. There has also been an improvement in the share of teachers who are providing feedback to pupils and their parents on their pupils' academic progress.

**Actual instructional time in Standards 1 and 2 increased dramatically between baseline and endline**, largely due to the change in curriculum but also due to a reduction in time lost. Compared with baseline, pupils are receiving 56 more minutes (maths) and 203 more minutes (Kiswahili) of instructional time per week on average. The introduction of the 3Rs curriculum increased the expected hours for these subjects, substantially in Kiswahili. At endline, more time is lost than at baseline as a proportion of official intended time due to it not being scheduled, but less time is lost due to teacher classroom absenteeism. Pupils are still only receiving around 60% of the intended instructional time.

**The evaluation finds some positive changes in indicators of gender-responsive teaching.** The gender balance of teachers' interactions with pupils in the classroom has improved significantly since baseline but is unchanged since midline. Teachers report that they have learned to be more aware and balanced on issues of gender, including seating arrangements, their interactions with boys and girls, and pupil roles in school chores.

Other objectives of the programme are to increase the inclusiveness of teaching for children. An area of improvement according to teachers interviewed in the qualitative endline is in addressing different learning levels among pupils within the same class. One teacher, for example, mentioned using repetition to help 'slow readers'. Another teacher described using streams within a class to separate out pupils at different performance levels and ensure a more individualised approach to their needs. The change in the spatial inclusion of teachers' interactions with pupils in the classroom has been mixed. On one hand, more teachers were engaging with at least one pupil from all six areas of the classroom<sup>41</sup> than at baseline. On the other hand, the proportion of teacher interactions with pupils in the back two areas of the classroom has decreased significantly since both baseline and midline.

Despite a programme objective to reduce/eliminate the use of corporal punishment in schools, it is almost universal. Of all interviewed parents of Standard 3 pupils, 96% report that the school practises corporal punishment, whilst 98% of these report that their child was beaten at school as a punishment in 2017. Although corporal punishment is allowed in Tanzanian schools, all three rounds of qualitative research have found that corporal punishment is a source of conflict between parents and schools.

For more information see EL Quant Vol. 1, Section; 4.5, 4.6, 4.7, and 7.3; and EL Qual, Section 5.1.

## 5.2 What factors explain the results?

Whilst the findings at the outcome level are not a straightforward positive story, it is possible these measures of teaching practices do not capture some changes in the quality of teaching, which have contributed to the programme's impact on pupil learning outcomes. The following sections discuss the extent to which the inputs and outputs were achieved and the factors which may have affected the links from inputs through to outcomes.

---

<sup>41</sup> For this indicator, the enumerator divided the pupils in the classroom into two columns and three rows based on where they were sitting.

### 5.2.1 Were inputs (residential and ward/cluster training, TLMs) delivered as intended?

**Provision of the teacher in-service training was fairly comprehensive at covering all schools but less comprehensive at the individual teacher and module level, and TLMs generally reached schools.** Thus, the programme did largely deliver the inputs for sub-component 1A based on its adapted design (but many activities were dropped from the original design).

The away-from-school training has been implemented as intended, with all programme schools having had at least one teacher attend sessions with all modules covered up to the time of the quantitative endline. Almost all Standards 1 and 2 teachers (98%) had attended EQUIP-T in-service training in 2016 or 2017, and 71% had attended both training away from school and school-based training sessions. Although implementation of the school-based sessions has been varied, 85% of schools report having completed all the modules, which may be considered a substantial success.

On the other hand, the ward cluster reflection meetings have not taken place entirely as intended. Almost a quarter of schools 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 meetings in 2016–17 (there should have been eight days – one per quarter).

EQUIP-T's TLMs have generally reached schools, but not all schools have received all of the materials. Almost all schools report receiving 'big books', around 90% report receiving literacy toolkits and supplementary readers, but only 71% of schools report receiving the numeracy toolkits and only 36% of schools say they have received the teacher 'read-aloud' books. The posters on a positive and safe learning environment were distributed later than the TLMs, and only 26% of schools report receiving these in 2016 or 2017.

For more information see EL Quant Vol. 1, Sections 4.2, 4.6, and 7.3.

### 5.2.2 Were outputs of the teacher component achieved as intended? If not, why not?

The sub-component largely did deliver its planned outputs as set out in the adapted design.

**The first output is the operation of the COL structure, specifically the school-based in-service training** – whilst EQUIP-T directly provided the away-from school training, the running of school-based training depends on school staff to organise the sessions. In the quantitative endline survey, 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. There is, however, wide variation in the implementation across programme schools: in 2017, 45% of schools held roughly the expected number of sessions, with the rest holding fewer (though this had improved since 2015). Furthermore, the proportion of teachers who have completed all the modules is lower than the proportion of schools: 46% have completed the Kiswahili modules, 58% the numeracy, and 72% the gender module; indicating that teachers do not attend all sessions. In addition, School Performance Management Meetings (SPMMs), another element of the COL structure at school level, are taking place in most schools but not always on a weekly basis and for their intended purpose.

The component's results chain assumes that early grade teachers attend the school-based training sessions in order to benefit from the content and shared learning. As discussed above, although all Standards 1 and 2 teachers had attended some EQUIP-T in-service training, many early grade teachers are not regularly attending the school-based training, and on average Standards 1 and 2 teachers report attending less than half the number of days of school-based training held. Schools find it difficult to schedule these sessions without having to cancel classes or shorten the school day – a

barrier which was identified by the baseline TOC analysis – and with high workloads, teachers struggled to engage with the sessions even though they want to.

Almost all schools have an in-service training coordinator; however, turnover is very high, which is particularly problematic given the coordinator's role in organising and facilitating school-based training, as well as the relationship required for effective support to other teachers. WEOs and head teachers see the in-service coordinator as essential for organising and maintaining school-based sessions.

More teachers reported challenges with the training in the quantitative endline than at midline. For training away from school, the most often cited were: limited training time; an insufficient allowance; and not enough training materials. For in-school training, the challenges reported were: lack of an allowance; participants not being motivated; limited training time; not enough training materials; and sessions taking place at an inconvenient time. Similar feedback emerged from the qualitative study, with a clear preference for away-from-school training. This is perceived as superior due to the access to a professional tutor, allowances, and scheduling; these aspects acknowledge that teachers are taken away from normal work. However, duration was still an issue, with the away-from-school training considered insufficient for these attendees to be confident of passing what they had learned to their colleagues back at the school. These teachers then had to navigate tensions with their colleagues given that they have received allowances for attending the away-from-school sessions. On balance, though, teachers were enthusiastic about the training they had attended. Materials were considered a challenge: many teachers in a school had to share a small number of modules; and TLMs were apparently delivered late.

**The in-service training was intended to improve teachers' capacity and confidence, and teachers report that it has done so.** 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%). The qualitative endline findings resonate with those from the quantitative survey and go further in explaining the benefit teachers now feel from having a mix of tools and techniques they can use to teach a topic and manage the classroom. New techniques, such as singing, group work, story-telling, and question and answer, means they have different ways to get a topic across and help pupils learn, as well as focus attention and manage class behaviour. The baseline qualitative research found that teachers lacked knowledge of the curriculum – what should be taught to a group of pupils, what the national syllabus and examinations contained, what textbooks to use, and how the school and grade level planning documents worked. Although the baseline quantitative survey did not measure this, the midline and endline surveys found that teachers' confidence in teaching the new Standards 1 and 2 curricula is high (only 1% are not confident).

There was also success in increasing the availability of TLMs in classrooms: the proportion of observed lessons with instructional materials displayed on the walls increased substantially between baseline and midline, and further by endline. However, the overall availability in classrooms of TLMs provided by EQUIP-T is low, with supplementary readers available in only 14% of observed Standard 2 Kiswahili lessons at endline, despite the majority of schools having received them. Whilst the reason for the lack of use of these TLMs was not explored, the failure to use them fully creates a substantial barrier to this intervention having the intended effect.

For more information see EL Quant Vol. 1, Sections 4.2, 4.3, 4.6, and 5.3, and EL Qual, Sections 5.1, 5.3, and 5.4.

### 5.2.3 Did the assumptions required for outputs to lead to outcomes for the teacher component hold?

**The improvement in teacher classroom attendance and instructional time is positive for ensuring that the intended benefits from training and use of materials can be put into practice, but absence is still high, as is teacher turnover.**

High teacher school attendance and instructional time are necessary conditions for the in-service training to benefit children's learning. In theory, the activities of EQUIP-T could overcome challenges of low instructional time by means of increased intrinsic motivation from greater job satisfaction or extrinsic motivation from greater accountability and monitoring by supervisors. The most notable change is an improvement in the level of classroom absenteeism for Standard 1 and 2 teachers: at endline, 35% of teachers are not present in classrooms despite being present at school and timetabled to teach, and this has dropped significantly, by 40%, since baseline. The large workload is the most common reason reported by teachers for their absence from the classroom. There has also been a significant improvement in teachers' punctuality: the share of teachers arriving late at school is still high, at 40%, but this is a reduction (from 63%) at baseline. However, looking across teachers of all standards, on average 13% of teachers were absent from school (unchanged since baseline) and classroom absenteeism for teachers of all standards is extremely high, at 61% (also unchanged since baseline).

Teacher turnover is very high: 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. This affects the effective delivery of the in-service training, particularly the reliance on a school-based, in-service training model that is continuous and sustainable.

**Teachers have faced worsening conditions, such as increasing class sizes and workload. This has affected their ability to put the learning from training into practice. The wider working environment has remained largely the same.**

Class sizes have been markedly increasing since baseline, reaching critically high levels at endline. The average class size for Standard 1 grew significantly, from 75 pupils at baseline to 104 pupils at endline; similarly, the average Standard 2 class size is 109 pupils, an increase of 68% since baseline. Large class sizes, and the associated workload, affect teachers' motivation, their ability to prepare, deliver, and follow up on lessons, and the techniques they can use in the classroom. Clearly these issues are very serious and EQUIP-T could only hope to mitigate these in part; teachers feel that some of the in-service training content was not appropriate for such large classes.

Whilst workload is not directly measured, a number of indicators point to an increase in this for teachers. At endline, they are preparing and marking more pupil assessments, for much larger classes, and they are also participating in more meetings due to the introduction of the COL. Overall, this contributes to classroom absenteeism and may lead to less time for lesson planning.

Challenges in teachers' working environment affect their motivation as well as their attendance and punctuality, and so could be a barrier to the component's effectiveness. The availability of nearby housing emerged prominently in the qualitative study, with long distances to walk to work, making teachers tired and demotivated, with a consequent impact on their teaching. The proportion of schools without a single teacher house (11–13%) and the average time for teachers to get to school (15–18 minutes) remain largely unchanged since midline. There has been a slight improvement in teachers' reported job satisfaction since baseline; however, when asked how their level of job satisfaction compares to two years ago, there has been a negative change since midline. Another underlying

factor for job satisfaction is that a majority of teachers are still paid non-salary allowances late. This has not changed since midline.

For more information, see EL Quant Vol. 1, Section, 4.4, 4.5, and 4.7 and EL Qual, Section 5.2.

#### 5.2.4 What unanticipated and unavoidable factors affected the results?

**The implementation of EQUIP-T's sub-component 1A, and its results, were affected by a number of unforeseen external factors, particularly the revised curriculum, the fee-free education policy, and the government's focus on work ethic, which had different effects on the results.** As these were all national changes, they would contribute to changes across the country, and in some cases prompted adaptations to the design.

The introduction of a revised curriculum in 2015 was complementary to the training delivered by EQUIP-T. The competency-based curriculum for Standards 1 and 2 focuses on the 3Rs, rather than a larger set of subjects. It promotes a phonics-based approach to teaching children to read. EQUIP-T's focus on literacy and numeracy is complementary, and thus the two reinforced each other – the EQUIP-T modules became more relevant for teachers in the light of the new curriculum, and the modules developed after the curriculum changed were properly aligned. The new curriculum markedly increased the prescribed number of instructional hours per week for Kiswahili in particular, but also maths, and would thus contribute to an increase (nationally) in focus on literacy and numeracy. Overall, through increasing the relevance and increasing instructional hours on literacy and numeracy, the curriculum change was a supporting factor for achieving EQUIP-T's results.

The fee-free basic education policy, announced in 2015, led to a substantial increase in enrolment, causing larger class sizes, an increase in pupils from poorer households, and higher workloads for teachers as the supply of teachers did not match the enrolment. As an indicator of the vastly increased demand for schooling, mean enrolment in Standard 3 grew from 65 to 95 pupils from baseline to endline, a strongly significant change (nearly 50%). The increase in class sizes over the period suggests that the system has not been able to increase basic resources (teachers and classrooms) to cope. As large class sizes affect teachers' methods and workloads, the fee-free policy is likely to have been an inhibiting factor on EQUIP-T's success.

One effect of the class sizes is the decline in the use of many positive teaching practices measured in this evaluation. Large class sizes affect what practices teachers can use: in observed lessons with more than 80 pupils, teachers were less likely to use paired or group work, instructional aids, or 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. The new 3Rs curriculum reorganised lesson sequence such that lessons are now taught consecutively. As a result, teachers would have changed the way they use some of the measured practices (at the start and end of a lesson).

The final unforeseen factor was the work ethic introduced by the new government from late 2015. The government introduced the slogan *hapa kazi tu*. In the midline qualitative work, respondents credited this with influencing teachers' work ethic, and the endline qualitative district study found an emphasis on how the government regime had increased accountability to perform at all levels of the public system. This work ethic was a supporting factor in EQUIP-T's success.

For more information, see EL Quant Vol. 1, Sections 3.2 and 4.5.

## 5.3 To what extent are factors favouring the sustainability of the teacher professional development model in place?

### 5.3.1 Was the teacher component considered effective? Are key stakeholders committed to continuing these activities and practices? Is the capacity in place to continue the activities?

**Many factors for the sustainability of the teacher in-service training are in place: the component is considered effective at school and district level, and there is sufficient organisational capacity; however, the current national commitment and leadership will need to continue to pass a consistent message to school level, and the issue of allowances may become a sticking point.**

At the school level, the in-service training is considered effective, and although this topic was not the focus of interviews at LGA level, brief references (and the midline study) suggest that district-level officers perceive the training to have been effective at improving teaching and pupil learning outcomes. This is evident in teachers' claims that they are already using and will continue to use the techniques learned that they find useful, driven by the successful impact on pupil understanding they report to be seeing already. However, the next stage for sustainability would be for the school-based training sessions to continue. Schools report that these sessions can and will continue with the head teacher's support and their integration into routine school meetings. The in-service coordinator is seen as integral in sustaining training at the school level, and ensuring school-level meetings continue. Whilst the commitment is there at the school level, the in-service coordinator position was created by EQUIP-T and thus is at risk of no longer being recognised after the programme ends.

Teacher turnover is both a risk to sustainability and a reason why sustainability is so important. Turnover is very high in these schools: as reported, 47% of Standards 1 to 3 teachers from baseline are no longer at the same school at endline, and in 2018, only 43% of schools had the same in-service coordinator in post since January 2015, when school-based training started. This turnover means that the relationships and culture of COLs may be disrupted; in particular, the teachers who have received the professional training away from school may leave (though they may continue in another school). This is especially problematic if the in-service coordinator leaves. On the other hand, this turnover is exactly why continuing the school-based sessions is seen as valuable: it makes sure new teachers also benefit from the content and approaches.

Schools want to see commitment and support from various levels of government in order for the school-based training to be sustainable. The in-service coordinator role needs to be formally recognised in the government system. Government needs to emphasise, or even mandate, these sessions, and support and monitoring must take place from central, through regional, LGA and WEO levels, and could be integrated in the SQA system. The government has adopted the in-service training as one of six 'best practices' that it wishes to continue after EQUIP-T ends, and the 2019 NF-TCPD (MoEST, 2019) includes a school-based COL model. This evaluation understands that MoEST is considering how to use the modules and the school-based learning model in a new national programme funded by GPE (LANES 2).

The human resource institutional framework does not currently incentivise teachers to attend school-based sessions. In-service training is not recognised in the responsibility and reward system, although the new NF-TCPD states the intention to change this, which is a positive sign. The lack of allowances and low motivation for attending school-based training is a sustainability challenge. Schools would like to see LGAs cover the cost of at least refreshments. The need for this is questionable given that these

sessions do take place in any case, but incentives would most likely increase attendance and the drive to keep holding school-based training and improve how the training is received.

The organisational capacity is largely there to continue the in-service training model. The modules were developed closely with TIE, and the national teacher training system was used to roll this out: lecturers from the public universities, tutors from the TTCs, and then WEOs and school teachers are each trainees and trainers in turn. WEOs say they have learned about what teachers are meant to be teaching and are now in a position to support and monitor this.

For more information, see EL Quant Vol 1, Sections 4.2 and 4.4 and EL Qual, Section 5.4.

### 5.3.2 Is it affordable for the government to continue to replicate and scale up the in-service training model?

**Continuing and scaling up the in-service training model will have a substantial cost – on average, 12 to 16% of the current non-salary budget allocated to LGAs for education – and this is unlikely to be affordable in existing budgets unless a substantial effort is made to increase education funding from the national budget.**

The endline cost study estimated the costs of continuing the in-service training in nine existing regions and scaling-up the modules (13 on literacy and on 13 numeracy, plus ones on gender-responsive pedagogy and general effective pedagogy) to the other 17 regions of the country. The analysis included a maximum scenario, which involves holding district-level refresher training every three years, and a minimum scenario which relies on school-based and ward/cluster level training after the initial roll-out.

Whilst costs vary substantially between years depending on whether district-level residential (initial or refresher) training is being held, the average annual cost over five years in a maximum scenario is TZS 100 billion (approximately £33 million) per year. Of this, TZS 30 billion per year is the cost of salaries for the time spent by public officials organising or attending training. If this salary cost is assumed to be absorbed within existing capacity, it leaves TZS 70 billion which needs to be budgeted for. Over half of this amount is for paying allowances to the participants and facilitators for attending various training sessions (based on government standard allowance rates). Other relatively smaller costs are for transport allowances (for attending regional, LGA, and ward/cluster sessions), and venue and stationery costs. The cost of purchasing and distributing the training manuals and modules to new regions makes only a very small contribution to the overall cost. There are also fees paid to university lecturers for training TTC tutors, and to TTC tutors (for LGA training and ward/cluster training). Over half of total costs comes from the district-level residential training.

In the minimum scenario, the only district-level training is for initial roll-out of modules to new regions. Then, the average annual cost when this is spread over five years is TZS 75 billion. Excluding salaries, this is TZS 50 billion which would need to be budgeted for.

To put this in context, in 2018/19 the total non-salary budget to education in the LGAs was TZS 422 billion. Assuming the salary costs are absorbed by existing staff time, only the non-salary costs are additional. In the maximum scenario, the annual average non-salary costs are 16% of total LGA non-salary budget, and in the minimum scenario, almost 12%.

LGAs have very little room for discretion in spending their non-salary budgets. The development budget is automatically allocated and directly transferred for *elimu bure* payments (the government's payments to schools to enable education to be provided to students without charging fees) by MoFP to schools. The 'other charges' (OC) budget (the budget for non-salary recurrent spending) is usually

absorbed by teachers' allowances and employee benefits. LGAs generally report that they do not have funds for routine monitoring, so organising and paying for training is an unrealistic expectation. It is not possible that LGAs would be able to fund these activities within the status quo budgets they receive.

This leaves the question of whether the government might increase the allocation to the education sector to cover these costs. Whilst general government revenue, and budget, is forecasted to increase in real terms in the coming five years, the education budget has been falling in nominal and real terms in recent years, which suggests that other sectors or commitments are currently a higher priority for the government. Furthermore, there will be an increase in costs in the education sector just to maintain current standards, with increasing enrolment requiring more teachers, more *elimu bure* payments, more infrastructure. In addition, in interviews for the qualitative study, LGAs report that their total own source revenue budget is being hit by changes made by national government to local business taxes,<sup>42</sup> which means LGAs receive less. Thus, if LGAs' discretionary own source revenue is reducing, it is unlikely that education will receive an increase from the LGA budget.

Based on recent trends, it does not appear likely that the education sector will receive substantial additional budget, even if the government as a whole has more fiscal room. The sector should make the case for more budget but with greater understanding of what the current government priorities are.

For more information, see EL Cost, Sections 4 and 5.

## 5.4 Teacher component summary

**Table 4. Summary of findings for the teacher component**

### Was the teacher component effective at achieving its objectives, and why?

**It has been partly effective: the component has increased the capacity and motivation of teachers, although exactly how that has fed through to the improvements in learning outcomes is unclear. The component was not effective, however, in increasing the use of reading materials in lessons.**

The results on the intended outcomes of positive teaching practices are mixed. Based on the measures used in the quantitative lesson observations, some have improved and others have significantly worsened.

Teachers' spatial and gender inclusion of pupils in the lesson has improved. Although it is not a clear-cut positive result based on the measures in the survey, teachers report that they have gained new skills and increased confidence from the EQUIP-T in-service training and that they are putting the tools into practice, such as classroom management techniques, singing, and story-telling. Some practices are difficult to use given the large class sizes, but at the same time the results are impressive given the substantial increase in class sizes driven by the fee-free education policy.

The inputs and outputs for the in-service training were largely delivered as intended, with almost all teachers attending some in-service training. However, there is still wide variation in the implementation of the school-based training, with some schools holding far fewer sessions than expected, and targeted teachers appear to be attending less than half of the sessions. Challenges include high workloads, lack of room in the school day, and low motivation due to lack of allowances.

Teachers are using different instructional materials more frequently in their lessons. However, despite the 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

<sup>42</sup> This is due to two changes: (1) a property tax used to be collected by LGAs but since October 2018 is now being collected by the central government; and (2) the central government has abolished and/or reduced tax rates and levies collected by LGAs.

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.

Actual instructional time in Standards 1 and 2 increased dramatically between baseline and endline, partly due to an increase in teacher attendance in the classroom; this may be related to higher motivation, but is more likely due to the change in curriculum and the allocation of more hours to literacy and numeracy.

### Are factors favouring sustainability in place?

**Most of the factors are in place; however, the in-service training model is not affordable within the current LGA education budget and additional resources will be needed.**

Many factors for the sustainability of the teacher in-service training are in place: the training is considered effective at school and district level, and the organisational capacity is in place; however, the current national commitment and leadership will need to continue and pass a consistent message to school level, and the issue of allowances may become a sticking point.

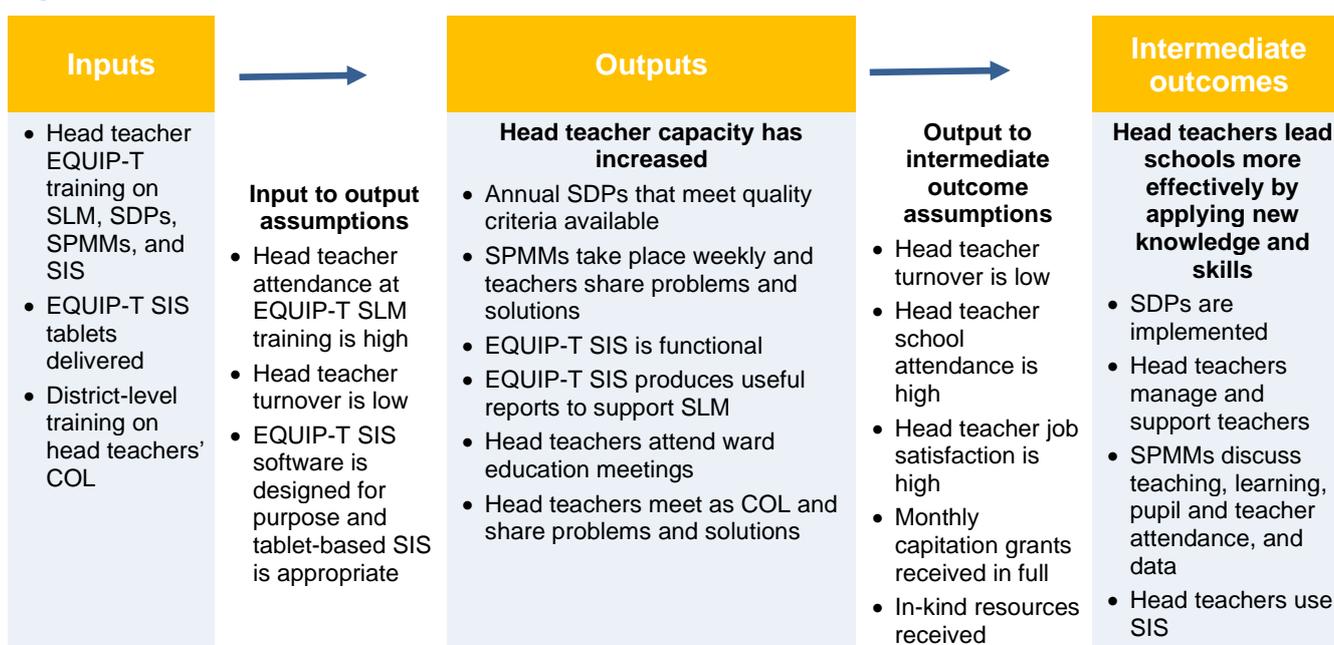
Continuing and scaling-up the in-service training model will come with a substantial cost relative to LGAs' resources – on average 12–16% of the current non-salary budget allocated to LGAs for education. Although not necessarily an expensive intervention for this setting, it is unlikely to be affordable if a substantial effort is not made to increase education funding from the national budget.

## 6 SLM component

The findings on component 2 – SLM – focus on the question of effectiveness, and touch briefly on sustainability. This chapter begins with the evaluation results chain (a sub-set of the TOC), which was used for assessing effectiveness and sustainability.

Component 2’s inputs include training for head teachers on SLM, which included SDPs, setting up SPMMs, and using the SIS. The SIS involves distributing tablets, which schools use to enter data regularly about pupils, teachers, and the school environment. Inputs also include encouragement of COLs at ward level, where head teachers come together to share experiences and learning. The inputs are expected to lead to outputs, including schools having SDPs, holding SPMMs, maintaining a useful, functioning SIS, and head teachers attending COLs with other head teachers in the ward. These outputs lead to more effective school management – managing teachers’ performance, implementing SDPs, using the SIS, and using SPMMs to improve school performance. Key assumptions in this results chain relate to the low turnover of head teachers, high attendance at training and in school, functioning of the SIS software and hardware, high job satisfaction, and receipt of resources for implementing SDPs.

**Figure 13. Main results chain for EQUIP-T component 2, SLM**



EQUIP-T’s logframe indicators which relate to this results chain are: Output indicators: (1) Percentage of head teachers achieving management performance indicators; (2) Percentage of Head Teachers attending COL meetings at least four times biannually; (3) Number of schools with at least 80% of enrolled students’ baseline data already uploaded into the SIS; and (4) Number of teachers and LGAs staffs trained in use of SIS.

The full evaluation framework is included in Annex B, with component 2 questions in Table 11.

### 6.1 Did the SLM component achieve its outcome targets for SLM?

**This evaluation finds that there has not been any significant overall improvement in head teachers’ SLM practices since baseline.** Some practices have improved; however, others remain unchanged or have even worsened.

In terms of managing teachers, head teachers’ use of performance appraisals has not changed significantly over time, and remains low, at around 30%. Significantly fewer lessons are observed, by

head teachers or other school staff, than at midline, and written feedback following observation is rare. However, improvements include a significant increase (from 91% at baseline to 96% at endline) of Standards 1 to 3 teachers who report that the head teacher checks their lesson plans; it is significantly more common to have rewards for teachers who perform well (55%) than at baseline (34%), according to head teachers. The large majority of head teachers (80%) report that action is taken if teachers perform poorly (this has not changed), and teachers broadly agree. Most head teachers consider teacher attendance at their school to be 'good' or 'very good' (92% at endline), whilst actual teacher absenteeism was 13%. More than half think it has improved over the last two years, even though the measured teacher absence shows virtually no change.

In terms of broader school management, there has been a significant increase in regular staff meetings between the head teacher and teachers 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. All head teachers report taking action to improve education in the school in the previous school year, similar to the finding at midline. Reported action to increase pupil attendance has risen significantly from midline to endline, from 17% to 45%, although levels of pupil attendance did not improve over the same period.

The development and use of SDPs to guide school management and increase transparency is a key intended outcome. At endline, only 49% of schools have an SDP and 41% have started implementing it, suggesting that once schools have developed an SDP, they use it, and that the challenge is to encourage schools to prepare SDPs to begin with.

**At the time of the quantitative endline, the SIS was not providing useful reports to support SLM, largely because a majority of schools had 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 had replaced other records and reports in 24% of these schools with functional tablets. At the time, there were a range of reasons for this poor result, including poor internet connectivity in schools, insufficient training of head teachers, and design flaws in the software. At least some of these constraints could have been anticipated. The EQUIP-T MA reports that since the survey, PO-RALG and the SIS application developer have worked to develop a more stable version of the application, one that is also able to generate a larger number of reports; this was finalised in October 2018. This improvement may have helped alleviate some of the problems with data entry and use of the SIS identified by the survey.

**The programme has exceeded its logframe targets for SLM in 2019.** The logframe has an indicator which measures a composite of school management factors. Measured using the annual monitoring survey (administered by WEOs in 2019), the indicator looks at the proportion of head teachers demonstrating a certain number of 10 performance indicators (such as SDP implementation started, weekly management meetings with teachers, and follow-up on teacher attendance from SIS). This indicator was introduced in 2017 and reported in 2017, 2018, and 2019. In 2019, the programme had exceeded its targets: in the original seven regions, 74% of head teachers achieve at least 8 out of 10 performance indicators, against a target of 60%. In the two new regions, 97% achieve 6 out of 10, against a target of 60%. This evaluation has not verified the logframe indicators. Other logframe targets for component 2 come under output and input levels, discussed below.

For more information see EL Quant Vol. 1, Sections 5.4 and 5.7, as well as Annex C.

## 6.2 What factors explain the results?

The following sections discuss the extent to which the inputs and outputs were achieved and the factors which may have affected the links from inputs through to outcomes.

### 6.2.1 Were inputs (SLM training, SIS tablets, and training) delivered as intended?

**Most head teachers had received some type of SLM training, but coverage was not complete and did not include all the topics originally intended.**

By endline, 76% of head teachers had attended some type of SLM training during the last two years, a significant increase from only 11% at baseline (and 68% at midline). However, 20% of head teachers had still never attended any EQUIP-T SLM training, which will likely undermine EQUIP-T's impact on SLM. The EQUIP-T SLM training has covered the intended topics, but most head teachers have not received training on all the topics. Although EQUIP-T developed a Leadership Competency Framework before midline as a basis for performance appraisal, implementation stalled, and the approach was later superseded by materials developed together with the Agency for the Development of Education Management (ADEM); thus, head teachers had not received training on using performance appraisal.

Furthermore, those who attended training attended fewer days than expected. Possible reasons for this are that some head teachers sent teachers to attend SLM training events on their behalf, and that head teachers missed some of the training events.

Head teachers who attended EQUIP-T SLM training report difficulties which are almost completely related to the design, arrangement, and location of the training. At endline, the most common reported difficulties are: no/insufficient payment (25%); limited training time (25%); envy from colleagues (15%); too much content covered (15%); sessions at/on an inconvenient time/day (14%); not enough training materials (10%); and transport difficult or venue too far away (8%).

In addition to SLM training, at endline, the vast majority (89%) of head teachers have attended some kind of early-grade teacher training, compared to none at baseline. This shows very high delivery of this activity, but the 11% who did not receive the early-grade teacher training may not be able to provide the intended academic leadership and teaching support.

**Less than half of head teachers had received the intended SIS training.** All head teachers were meant to receive training on SIS management after midline but the 2018 evaluation survey found only 43% of current head teachers did – this may be related to high turnover (discussed in the next section). The logframe has an input target for the number of educational managers who had received training on SIS data use. In 2018, this target was met and 6,519 had received training, according to data from the SIS database and District Executive Director report, in the logframe. However, the logframe reports that the 2019 target was not met, with only 5,196 trained –1,300 fewer than the number who had been trained the year before. It is not clear if this is meant to be a cumulative total, in which case it should not decrease year on year. The logframe indicator also presents the challenge with relying on absolute numbers of trainees: a large number of participants may have received training but there is not high coverage of schools.

Almost all schools (97%) had received the SIS tablet before the endline survey, and almost all of these tablets are functioning. Tablets were charged in only 55% of schools, despite solar chargers being supplied with the tablets.

For more information, see EL Quant Vol. 1, Sections 5.3 and 5.7, as well as Annex C.

## 6.2.2 Were outputs of the SLM component achieved as intended? If not, why?

**The SLM training included preparation and implementation of a SDP, and there has been a notable increase in the proportion of schools with an SDP**, from 36% at baseline to 49% at endline; however, this is a drop from 68% at midline. The increase between baseline and midline may reflect the fact that the initial training on SDPs was delivered around the time of the midline survey. On average, SDPs contain more of the core elements at endline. Compared with 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%). Only 2% of SDPs at endline contain baseline data and targets, a decline since midline, and this is at odds with EQUIP-T's aim of promoting the use of data for school decision making and to improve SLM.

**SPMMs, meetings led by teachers but established following training for head teachers, are taking place in most schools.** However, 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, as intended; the discussion topics suggest they are not always being used for their intended purpose.

**The substantial level of head teacher turnover undermines the extent to which leadership capacity can improve as an output of the SLM training.** Only 26% of head teachers who were at the school at baseline were still at that same school by the endline four years later. Among the 24% of head teachers who did not receive SLM training in the last two years, the reason for all but one was that they had joined the school since the end of 2017, after the training had been conducted. At endline, the majority (91%) of recently appointed head teachers 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. Though turnover was already high between baseline and midline, it was exacerbated between midline and endline by the change to the minimum qualification policy, discussed in Section 6.2.4.

**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: increased awareness of head teacher responsibilities (71%); improved teacher management (59%); better relationship with teachers (50%); more confidence in their role as head teacher (44%); and improved relationship with parents and the community (44%).

**At the time of the quantitative endline, there was very little evidence that the SIS was being used and found useful.** Only 24% of schools with functional tablets had entered up-to-date lists and data for both teachers and pupils in the school at the time of the endline. 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 (the fieldwork was three to four months into the year), whilst this is expected to be a tool used daily. **However, the programme exceeded its logframe target for SIS.** The target was for 2,500 schools to have at least 80% of enrolled students' baseline data already uploaded into the SIS in 2019. The programme reports that 3,689 schools have this data uploaded, far exceeding the target.

For more information, see EL Quant Vol. 1, Sections 4.2, 5.3, 5.4, 5.6, and 5.7, as well as Annex C.

### 6.2.3 Did the assumptions required for outputs to lead to outcomes for the SLM component hold?

**High head teacher absenteeism (21%) affects the ability of heads to translate their increased capacity into improved SLM practices.** The main reasons given for absence are official and work-related: attending meetings (76%) and official government work not related to education (29%). Head teachers' self-reported job satisfaction has remained unchanged since baseline.

**A number of design factors have hindered schools from using the SIS even when their tablets are functioning.** For schools that have functional tablets but no or incomplete records at the endline survey, the main reasons for having such records are the time taken to enter the data, the failure of tablets to work properly, and head teachers not understanding how to enter data into the SIS and/or not receiving training on this. These factors reflect issues with the design of the application software to make it easy and accessible, as well as the design of training, aspects which EQUIP-T could address. Head teachers also report difficulties which are harder for EQUIP-T to address: poor internet connectivity (79%); high workload and the 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 the lack of a regular electricity supply to charge the tablet (9%).

For more information, see EL Quant Vol. 1, Sections 5.5 and 5.7.

### 6.2.4 What unanticipated and unavoidable factors affected the results?

**Two national policy changes have affected the intended results chains for component 2 and thus the actual results.**

**The new national minimum qualification policy was implemented in 2017 and both raised head teacher qualification levels and caused substantial turnover.** The policy requires all head teachers to have a diploma, and as a result, head teachers have become significantly more qualified professionally since baseline. At endline, 5% 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 is likely to have a positive effect on achieving the programme's objectives if the higher qualifications translate into higher skills and capacity, including for leadership. On the other hand, the change likely contributed to the level of head teacher turnover, and whilst it may be a one-off event, it was a major disadvantage to the programme, relying as it did on capacity building.

**The government changed the mechanism for transferring capitation grants to schools in January 2016, and this has led to schools receiving more reliable payments, which is a factor supporting school development activities.** A key assumption in the programme TOC to facilitate better school development planning and implementation is that schools receive their capitation grants in full. There was a significant and large improvement in this indicator after the mechanism changed. Grants are now paid directly from MoFP to schools, rather than via LGAs; in 2015, none of the programme schools received the capitation grant in full, whereas in 2017 61% did. There has also been a significant increase in the proportion of schools receiving in-kind resources since baseline, particularly classroom furniture, in line with the government directive to increase the availability of desks in classrooms.

For more information, see EL Quant Vol. 1, Sections 5.2 and 5.6.

### 6.3 To what extent are factors favouring the sustainability of the SLM component in place?

Whilst the issue of sustainability was not explored in the supporting studies for the endline evaluation, the findings from the 2018 quantitative survey are not promising for the sustainability of the SLM activities. The high level of head teacher turnover indicates how quickly the value of the initial training is lost to that school, unless the new head teacher is also trained. The importance of this training is evident in the results on SDPs: the share of schools with SDPs increased soon after SDPs were covered in the SLM training, and then the share dropped. Thus, repeated training or some form of follow-up mechanism appears to be necessary to sustain the practices learned at the training.

At the time of the survey, the SIS was not being actively used – most schools with functioning tablets were not entering the data. The operation of SIS may have changed in the final 18 months of the programme, but based on the situation at the time data was collected, it is very unlikely that any benefit will be sustained from this activity.

### 6.4 SLM component summary

**Table 5. Summary of findings for the SLM component**

#### Was the SLM component effective at achieving its objectives, and why?

**No, this evaluation did not find evidence of more effective SLM or effective use of the SIS.**

The outcomes for head teachers' SLM capacity do not show a notable improvement. Whilst there are improvements in some practices, such as providing rewards for teachers who perform well and holding regular staff meetings, fewer lessons are being observed than at baseline and the use of appraisals remains low and unchanged. More schools have an SDP than at baseline but this proportion dropped after midline. The relatively low outcomes from this component are likely due to the high turnover of head teachers (exacerbated by the new qualification policy in 2017) and head teacher absenteeism. They also raise the question of the effectiveness of the training content.

The SIS is not generally being used by head teachers, indicating a failure of this outcome at the time of the endline survey. Although most schools had functioning tablets, only one quarter were entering data into the SIS. This is likely because fewer than half of head teachers had received the training (again a challenge of turnover), and head teachers felt that data entry takes too long and do not understand how to use the system.

#### Are factors favouring sustainability in place?

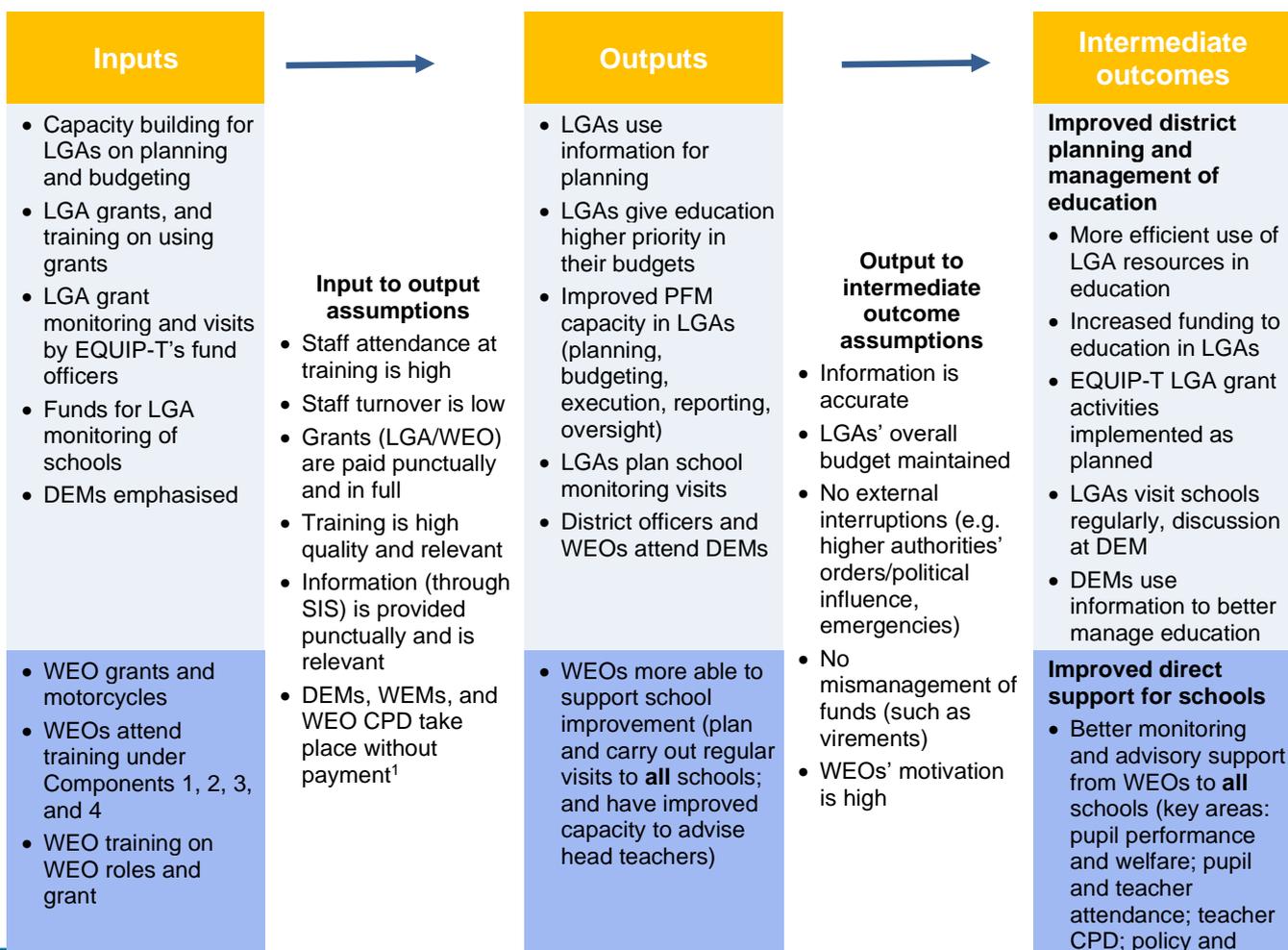
**Based on the evidence collected in the endline quantitative IE, it is very unlikely that a positive benefit will be sustained after the programme ends.**

## 7 District component

Component 3 aims to strengthen district planning and management and this affects two levels: the capacity and practices of district officers in the LGA offices; and the support given by WEOs to schools. This chapter addresses both of these, and answers questions on effectiveness and sustainability. The evidence is drawn largely from the qualitative study, which sampled three LGAs selected to represent best practice; such a sample, though small, should indicate the best possible results. This chapter briefly begins with a summary of the evaluation results chain.

Component 3’s inputs at the district office level are: capacity building on planning and budgeting; the provision of grants and training; monitoring and support on grant use; the provision of grants for monitoring schools; and the encouragement of DEMs. These are expected to lead to outputs: LGAs using information for planning and giving education higher priority in their budgets, having improved PFM capacity, planning school visits, and attending DEMs. The expected outcomes are: more efficient use of resources at the LGA; increased funding to education; the implementation of EQUIP-T grant activities; more school visits; and better management of education through DEMs. For WEOs, the inputs are motorbikes and grants, training under various components and specifically on WEOs’ roles, and the establishment of monthly WEO CPD sessions. Intended outputs are that WEOs attend CPD and have improved capacity to support schools, through their skills and ability to reach schools more frequently, with this leading to outcomes of improved monitoring and support to schools. Key assumptions for component 3’s results chains to hold are low staff turnover, high-quality training, COLs (DEMs and WEO CPD) taking place without payments, LGAs’ overall budget being maintained with no external interruptions, high staff motivation, and funds not being mismanaged.

Figure 14. Main results chain for EQUIP-T component 3



- WEO training on CPD (COL)

- WEOs attend monthly CPD meetings (COL)

curriculum; school finance and resources; school governance; community engagement)

Notes: (1) No payment except for transport, and a per diem if an overnight stay is required.

EQUIP-T's logframe indicators which relate to this results chain are: Output indicators: (1) Percentage of WEO attending CPD sessions at least twice in a quarter; (2) Percentage of LGA districts with complete reconciled financial reports for EQUIP-Tanzania activities from the most recent financial quarter; (3) Number of districts executing DEMs at least four times bi-annually.

In response to changing needs, this component has adapted substantially over the course of EQUIP-T's implementation. The original design, in which the MA was responsible for the largest share of programme funds, was soon considered an unrealistic way of managing the programme, and so from 2015 the design was changed to implement activities through decentralised funding to LGAs. This was also anticipated to increase local ownership and strengthen local PFM capacity. Whilst the programme adapted the amount of support it gave to LGA officers in response to the context (such as who was included in training, and by providing more on-the-job support), it also dropped some of the capacity building, which was intended to raise general PFM capacity in the education sector.

The full evaluation framework is included in Annex B, with component 3 questions in Table 12.

## 7.1 Did the district component achieve its outcome targets for district planning and management?

**There has been mixed success in achieving the intended outcomes for component 3. The intended outcomes at the level of WEOs have generally been achieved. At the LGA level, there has been success in introducing and using DEMs, but less success at improving planning and budgeting practices. This has likely come as a result of the need to focus on the implementation of LGA grants, and whilst this evaluation has not included a process evaluation, clearly many of the programme's activities have been successfully delivered through LGAs.**

**This evaluation does not find evidence of changed district planning and budgeting practices.** There are processes and products of LGA education planning but these are limited in the extent to which they guide implementation, given the unfamiliarity of most LGA level officers with the details of the plans. For example, LGA officers did not know the details of the examination targets, or the period covered by the LGA medium-term strategic plan.

**It is not clear that the funding to education has increased, and may even have decreased.** Various officers in one LGA visited for the study explain that the OC budget has fallen in recent years, with various explanations given: this is a result of the new government regime; it is due to decreasing general government revenue; or that the 2016 change from capitation grants to *elimu bure* payments shifted funds from LGAs so that they went directly to schools. This is validated by a recent fiduciary risk assessment. LGAs state that their own source revenues are falling as the government has changed local property and agricultural produce taxes, reducing the income LGAs can collect, and undermining an assumption that LGAs would continue to have the same resources available.

The markedly varied responses from LGA officers about the status of budgets signal a lack of awareness of budget details amongst LGA officers, and low transparency of budget and actual expenditure within LGAs – this is also confirmed by the fiduciary risk assessment in terms of lack of aggregate LGA budget information.

**LGAs have budgeted for EQUIP-T sustainability activities in their 2019/20 budgets.** This is an indication that LGAs may direct more own resources to education in order to continue EQUIP-T activities, although it is too soon to confirm whether this is happening in reality. Furthermore, it is not possible to identify and report total allocations to education from LGAs' own source budgets so it will be very difficult to verify these reports and confirm the total allocations and actual expenditure to the education sector in all LGAs.

**Accountability of LGAs to senior levels of government is perceived at endline to have increased over recent years, driven particularly by the current government regime.** EQUIP-T's monitoring of LGAs has also increased accountability, at least for EQUIP-T LGA grants, but this has not been institutionalised to improve wider accountability and ensure lasting change. EQUIP-T has required regular reporting from LGAs, with monthly reports submitted to regional secretariats. These EQUIP-T reports have made some LGA officers interviewed feel more focused on results, and pressured to complete activities. Meanwhile, according to EQUIP-T staff, the reports required for EQUIP-T have not been taken on systematically by the regional secretariats and PO-RALG.

**As a result of lower OC, officers report that the LGAs are making fewer monitoring visits to schools** – although some officers state that this is a deliberate shift towards more monitoring by WEOs, who now receive a responsibility allowance. LGAs have conducted school visits using EQUIP-T grants for monitoring. Although LGA officers advocate the importance of their own monitoring of schools, the qualitative study did not see evidence that monitoring is carried out systematically or that the information is used to direct action.

**DEMs are valued by all participants – education officers, WEOs, and SQAs – as bringing a number of benefits and improvements to the way education is managed in the LGA.** Many respondents see DEMs as a platform for learning from one another on how to resolve issues or on specific topics. The DEM provides a regular monitoring and accountability check-in for WEOs and as such is perceived by WEOs and LGA officers across the LGAs as making WEOs more committed. DEMs are seen by WEOs and education officers as an efficient way of receiving information from schools, and of raising and resolving issues quickly. The consultative nature of DEMs is felt to contribute to a better relationship between WEOs and other officers, as well as to giving WEOs more confidence.

**The programme has largely delivered through the decentralised implementation model.** With over half of all programme spending taking place through LGAs (and schools, for infrastructure), a substantial part of the programme's implementation has clearly been a result of the functioning of the decentralised funding mechanism. Whilst this evaluation does not include a process evaluation or a focus on fidelity of implementation, the evidence across the four components indicates that many activities have been delivered at the LGA level, which is a notable accomplishment.

When asked directly why some LGAs are higher performers, the answers from EQUIP-T staff tend to come down to the attributes of the LGA officers, and various aspects of leadership. Thus, these are not factors affected by EQUIP-T, and are not easy to replicate or influence through policies. Factors such as having a district education officer (DEO) with an open and collaborative style, close supervision of officers and schools, interest from the district executive director, and willingness to learn from key officers, were identified as important to success.

**Although this evaluation has not conducted an in-depth study at the school level to verify whether WEOs have improved, there are indications of improvements.** In the three LGAs visited for the qualitative endline, district-level respondents, and WEOs perceive WEOs as giving better support to schools, visiting schools more often, and generally being more motivated than a few years ago. **On the other hand, the quantitative endline found 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.** This change between 2016 and 2018 may be related to the substantial WEO turnover in that time (over half of schools had a new WEO between 2016 and 2018). The qualitative study may have had more positive findings due to the specific selection of high-performing districts and WEOs, and the bias from respondents.

The motorbikes and grants are seen to have led to closer monitoring and supervision, such that WEOs can more easily fulfil their duties and respond flexibly and quickly. WEOs are making more frequent visits to schools, and this closer supervision is felt by one head teacher and by education officers in all three LGAs to have contributed to teachers performing better, particularly by keeping teachers alert rather than allowing them to become lazy. It also means WEOs can attend meetings in the school locality, which improves the relationship with the community.

EQUIP-T staff and SQAs indicate that WEOs are communicating more effectively with teachers, acting more as a coach than as 'a policeman'.

For more information, see EL Qual, Sections 6.1 and 6.2, and EL Quant Vol. 1, Section 5.8.

## 7.2 What factors explain the results?

### 7.2.1 Were inputs (capacity building, monitoring, grants, motorbikes) delivered as intended?

**Aside from a shift away from formal capacity building, inputs have largely been delivered as planned.** At district level, LGA officers have received training and mentoring on education planning and management, but this has become more narrowly focused on EQUIP-T grant implementation and monitoring over time. LGAs have received the EQUIP-T grants, which are used across all components. WEOs have attended capacity building under various components and the initial training to conduct WEO CPD was delivered. WEOs have received motorbikes and WEO grants.

The inputs intended to strengthen district officers' capacity in planning, management, and more broadly PFM were a set of formal training sessions on education planning, training on managing the EQUIP-T LGA grants, and on-the-job support from EQUIP-T fund officers. The change in funding and implementation modality led to a change in delivery, not implementing the inputs as in the original design.

LGA officers were trained in modules in strategic and annual planning, and budgeting and budget management, in 2015 and 2016, as well as in budgeting and managing the decentralised grants which continued up to endline. However, there was less focus on providing inputs related to PFM than initially anticipated, as the programme had to focus on the infrastructure and support for implementing through decentralised LGA budgets, and meeting fiduciary responsibility for spending. The further modules originally planned were dropped as the MA felt they were not effective, and instead favoured more on-the-job support from EQUIP-T fund officers to support the implementation of EQUIP-T LGA grants – EQUIP-T refers to this as 'mentoring'. The qualitative study found that EQUIP-T MA regional staff and fund officers have a better understanding of fund implementation progress than general PFM functions in the LGAs, indicating their limited involvement in broader PFM. In addition, the focus on training in general education planning in the first two years was reflected in some positive responses about learning in the midline study. It is not surprising that there are so few specific changes reported at endline, given that there was little formal capacity building in education planning after midline.

LGAs have received EQUIP-T funds as per the LGA budgets, although there have been challenges in terms of delays in receiving grants and in implementation. Delayed transfers occurred due to the sequence of processes in central government. The perceived impact of these delays is unquantifiable: they can lead to clashes of timetables with local priority activities (such as examinations) and delays in experiencing the benefit; they can potentially have an impact on quality if, for example, training participants have to attend many sessions pushed into a short space of time. EQUIP-T staff also report that there have been delays in the intended workplan due to clashes with government activities such as examinations, which take priority.

The EQUIP-T MA has provided monitoring of EQUIP-T funds and activities, and regional and fund officers have provided feedback to help improve LGAs' reports.

WEOs report that they have received a number of inputs under EQUIP-T. In terms of capacity building, they have received training under all the components: on early grade literacy and numeracy under sub-component 1A, on SLM under component 2, and on the relationship with the community under component 4. Whilst there has been substantial turnover in WEOs over the last three years, WEOs had benefited from these training sessions (including in their previous positions as teachers). In all three LGAs visited, WEO CPD was initially rolled out with WEOs submitting forms to SQAs stating which topics they would like to receive training on. Topics were then selected and arranged in a timetable. However, in one of the three LGAs respondents confirmed that this has never been implemented. WEOs have received motorbikes and continued to receive the WEO grant to be spent on fuel, maintenance, and stationery.

For more information, see EL Qual, Sections 6.1 and 6.2.

## 7.2.2 Were outputs of the district component achieved as intended? If not, why not?

**At the district level, the functioning of DEMs is the most successful output achieved. District-level capacity improvement is not apparent. WEOs report that their capacity has increased and they are more able to visit schools regularly; WEO CPD is functioning to some extent.**

**The qualitative study did not find a notable strengthening of planning and budgeting capacity in the LGAs visited**, and these were selected as being relatively successful LGAs. LGA officers do not identify specific new knowledge or PFM skills that they have gained from EQUIP-T. In comparison, at midline, one DEO explained they had learned the 'grassroots planning process' from EQUIP-T training; few similar statements were made at endline, however. Furthermore, some LGA officers actually feel they have learned nothing new by using government systems. One planning officer said that a training he/she attended in 2015 on strategic planning offered nothing new since it had all been covered by previous training and professional experience.

Another reason that LGA officers have not gained capacity is that the EQUIP-T budgeting process is largely driven from the EQUIP-T MA and so does not allow LGA officers to practise the full suite of planning and budgeting skills. The process is top-down, with LGA officers only confirming the statistics used in budget formulae. Despite the intention to give LGAs more ownership with less centralised control, and hence more chance to practise planning and budgeting, this shift has not happened. However, some LGA officers see this budgeting process as efficient.

On the other hand, LGA planning officers and accountants (both groups outside the education department) report that through EQUIP-T they have been given greater exposure to education and now feel have a greater understanding of the needs. These officers speak of learning about the importance of community involvement, pupil welfare, the 3Rs, and the realities of education delivery.

EQUIP-T staff feel that LGAs have improved at filling in EQUIP-T's report templates over time: this relates to both monthly financial reports and activity reports. EQUIP-T staff feel that the support and feedback from fund officers and regional officers has contributed to this.

**In all three LGAs visited, organisation and attendance of DEMs is happening broadly as intended.** The main attendees are WEOs, education officers, and SQAs. The date and agenda of the meetings are organised by the WEOs' chairperson and secretary, in collaboration with a representative from the DEO's office, and whilst the LGAs intend to hold DEMs monthly, in reality they are less frequent. The agenda follows a similar structure: WEOs present performance reports for the last month, which include EQUIP-T activities such as IGA and PTPs, as well as other issues. WEOs present their successes and challenges, providing an opportunity for other attendees to give comments and suggestions for solutions. In two of the LGAs, the WEO update may be followed by a discussion of a specific topic identified as challenging, for which a facilitator is invited if the permanent attendees are not in a position to facilitate. This last session is likely the same thing as WEO CPD in some LGAs. DEMs are also captured in the programme logframe in terms of the number of districts holding DEMs at least four times biannually, and this target was almost met in 2019: 53 districts against a target of 55.<sup>43</sup>

Sharing of best practice at DEMs is seen by LGA staff and WEOs as a way that WEOs have improved their capacity. The routine presentation of reports in DEMs means that WEOs learn from their peers, and DEMs can be used specifically to train WEOs on writing reports. DEMs are also seen as the main source of information on WEO performance.

The logframe includes an output indicator relating to grant management: the proportion of districts with complete reconciled financial reports for EQUIP-T activities from the most recent financial quarter. For 2019, the target was 100% but the actual result was 0%, with a note that the delayed disbursement of tranche 7 meant that funds could not be spent in the intended financial year and it affected reconciliation. In this case the delay affected a PFM process but it is indicative of how delays may have affected implementation and achievement of the intended activity outputs. In previous years the target was met: over 97% of LGAs had reconciled financial reports in 2018, and the target was 95% for the old regions.

**WEOs report that they have a better understanding of their role and how to support schools.**

An overarching view is that whilst previously WEOs knew what their duties were, they had only prior experience as a head teacher to draw on rather than any instruction or training on actually how to fulfil their roles. WEOs feel that EQUIP-T, through routes including training and DEMs, has showed WEOs how to do their job. Particular changes in their understanding relate to putting more focus on their role in supporting community relationships, to administration and management matters, and to how to develop a workplan. According to district officers, WEOs have improved at writing reports over recent years, with a good report seen as one that has detail, is longer, and contains photos.

**WEOs are visiting schools more frequently than in the past.** The quantitative endline found that 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 significant and substantial positive change. Some WEOs say they can now spend a longer time in schools, some say they can visit more schools within a single day. WEOs report their visits to be of varied lengths – some spend a whole day assessing teachers in one school, others report visiting each school for one hour, or very briefly just to drop something off and count

---

<sup>43</sup> The logframe actually reports this indicator in two different ways: the number of districts holding DEMs at least four times biannually; and the number holding DEMs twice in a quarter. These are slightly different measures (one is quarterly, the other biannual), and it is not clear which measure the logframe is reporting against.

attendance. The quantitative survey found that 56% of the most recent visits to schools were less than one hour in duration.

**The ethos of a COL among WEOs, intended by WEO CPD, was in place in the sampled LGAs, but the implementation of WEO CPD as envisaged by EQUIP-T is inconsistent.** Two of the three LGAs visited were holding WEO CPD, though there was not a common understanding of what this is. Explanations varied: for some, it was a day conducted ahead of the DEM, or just a topic being discussed within the DEM; for others, CPD means a deeper topic for WEO training in the absence of LGA staff. In all cases, an appropriate facilitator is found for the technical matter. Training topics have included use of the SIS tablets, computer literacy, the performance appraisal system, academic camps, and managing school resources. SQAs do not appear to be as involved as EQUIP-T intended them to be. In the third LGA visited, after initial training on WEO CPD, the LGA did not start holding WEO CPD sessions. Thus, the assumption that the initial training would be enough to kick-start this was flawed.<sup>44</sup> The programme's 2019 logframe target for WEO CPD – the proportion of WEOs attending CPD sessions at least twice a quarter – was almost met for the original seven regions (79% against a target of 80%) and fully met for the two new regions (82% against a target of 80%).

The benefit of CPD is bundled together by WEOs with other capacity building and knowledge sharing that takes place through DEMs and more formal training. It is recognised as a way to improve WEOs' knowledge on areas they are inexperienced in, especially when they are new to the role; the demand-driven nature of the topics means it is relevant to WEOs' needs.

For more information, see EL Qual, Sections 6.1 and 6.2, EL Quant Vol. 1, Section 5.8, as well as Annex C.

### 7.2.3 Did the assumptions required for outputs to lead to outcomes for the district component hold?

**The evaluation identified a number of challenges with the underlying assumptions which prevent the intended outputs leading to the outcomes for the district component.**

Even if LGAs gained capacity in strategic and annual planning, they have little space for, and gain little value from, putting it into practice. Limited financial resources and directives from higher levels of government leave LGAs with little room to follow their own priorities. The OC budget is absorbed largely by teachers' leave and transfers, as well as necessary expenses such as teachers' funerals, leaving little money to be spent at the discretion of the LGA.

**Virements and high staff turnover have created barriers to effective implementation of the EQUIP-T LGA grants.** Virements take place when funds which were meant to be for EQUIP-T activities are used for something else (possibly another sector). EQUIP-T MA staff report that virements have been an issue in some LGAs;<sup>45</sup> fund officers have focused on resolving issues involving these, with government support, but understand they can take place when the LGA faces pressure to divert funds. In one LGA visited for the qualitative study, there were examples of funds being used for other activities; the money had already been returned for EQUIP-T activities. According to EQUIP-T staff, turnover within the LGAs, particularly of accountants and education officers, presents difficulties for the smooth implementation and application of learning given how key some of

<sup>44</sup> WEO CPD was not the focus of research in this LGA, so the reasons for its non-implementation were not explored.

<sup>45</sup> The quarterly report for Q22 identifies a total of eight virement cases reclassified as fraud cases.

these officers are. It requires more hands-on support from EQUIP-T staff to bring new officers up to speed.

**On the other hand, EQUIP-T, through the provision of motorbikes and WEO grants, and the government's WEO responsibility allowance, has contributed to WEO morale.** It has proved a contributing factor in WEOs being able to provide better support to schools, as well as increasing pressure on WEOs to fulfil their duties, as they now have 'no room for excuses' for not performing.

For more information, see EL Qual, Sections 6.1 and 6.2.

#### 7.2.4 What unanticipated and unavoidable factors affected the results?

##### **Numerous changes introduced by the government have affected this component.**

LGAs have less discretionary budget for education, for various reasons. The 2016 change of capitation grants to *elimu bure*, which is transferred directly to schools, means that LGAs no longer receive a development budget from central government. A government change to local property taxes and agricultural levies has reduced LGAs' ability to raise local revenue, so there is less budget to allocate to education.

Similarly, the government has also reduced LGAs' OC budgets and instead given WEOs a responsibility allowance, as of 2016, so shifting away from direct monitoring of schools by LGA officers to greater reliance on WEOs. The responsibility allowance means that WEOs have more resources to put towards fuel and food for lunch. This allowance has boosted WEOs' morale as they see it as a recognition of their position.

The government elected in late 2015 is felt to have increased pressure to fulfil public sector duties and deliver results, introducing much greater accountability at all levels of government. LGAs report being frequently asked for data. This is attributed to the higher levels of monitoring in place by the current government, meaning more requests from central government flowing down, but also to the increased pressure on intermediate levels (Regional and District Commissioners) to keep a closer check on LGAs. In turn, the LGA presses for more results from WEOs and schools.

Two techniques for WEO performance management appear to be more prominent at endline than in the past, and it is not clear what has contributed to this change. One is the Open Performance Review and Appraisal System, and the other is the use of ward guardians (each LGA education officer is responsible for overseeing a small number of wards). If these techniques are being effectively used, they would contribute to EQUIP-T's objectives of strengthening the management of WEOs in order to improve the support that WEOs give to schools.

The WEO professionalisation policy, announced in 2017, increased the qualification of WEOs but also led to high turnover. The policy requires that all WEOs should have a degree (rather than a certificate) by the end of September 2019. The two LGAs which had implemented the policy estimated that around 70% of WEOs were changed, and the higher qualification is seen as producing WEOs who write better reports, and are more effective and efficient. For the most part, district officers believe the new WEOs are performing better than those who were demoted. Whilst this change would support EQUIP-T's aims, this turnover is much more disruptive when looking at a broader scale: between the start of 2016 and 2018, 56% of schools have had a new WEO, according to head teacher reporting, which is a significant change over a short period of time. 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. Furthermore, new WEOs had not received previous rounds of training, and the programme used resources to help orient the new WEOs.

For more information, see EL Qual, Sections 6.1 and 6.2, and EL Quant Vol. 1, Section 5.8.

### 7.3 To what extent are factors favouring the sustainability of the district component in place?

Sustainability of the district component would mean the continuation of improved reporting and DEMs, and continued WEOs support to schools, which is more effective. For DEMs and the WEO interventions, these are generally seen as effective and there is commitment and knowledge at the local level to continue them. However, there needs to be commitment from higher government, with leadership and continued follow-up; lack of affordability may jeopardise the continuation of these activities.

Although broader than component 3, an intended level of sustainability under this component is actually that LGAs continue to implement activities across all the components for which they previously received EQUIP-T grants. Whilst LGA officials were positive about various EQUIP-T activities and keen for them to continue, it is not clear that all the factors for sustainability are in place. LGAs had been asked to budget to continue EQUIP-T activities – choosing from the government's identified 'best practices' – in the 2019/20 budget. Some EQUIP-T officers noted that it was difficult to convince LGAs to budget for these activities because they were accustomed to external financing and had problems with own source funding. Corroborating this, the qualitative study found that LGA officers were expecting EQUIP-T to be extended again and so were not anticipating having to fund activities themselves. However, most LGAs have included a budget for sustaining a selection of EQUIP-T activities, funded from their own source revenues.

Respondents have doubts about what will happen in reality – raising questions about whether LGAs will raise and receive the funds to run these activities. There has been no indication that the central government is ring-fencing funds for EQUIP-T sustainability activities. Overall, it is unlikely that the full funds budgeted will be made available, and therefore activities would only be able to continue in a much-reduced form. This is disappointing but perhaps inevitable for a programme that introduces new activities with cost implications.

At the LGA level, DEMs are considered effective. LGA officers report that they would like them to continue, and they have the skills and knowledge to do so. LGAs and WEOs also see DEMs as almost zero-cost – there is a cost for WEOs to travel to the LGA office, and there is usually a contribution for lunch, which will become more difficult in a context of lower resources for WEOs. The extent to which DEMs will continue depends on the commitment and capacity of LGA officers themselves to continue them, as well as commitment from higher levels. Although this evaluation has not assessed commitment at national and regional levels, officers interviewed for the qualitative study explained that the regional governments have indicated that DEMs should continue and this means they will. As with many EQUIP-T activities, there will need to be sustained leadership and commitment to holding DEMs from senior positions in PO-RALG, which is at risk with the frequent turnover of key government officials.

At higher levels of government, the lack of ownership of the EQUIP-T reporting system makes various EQUIP-T staff uncertain that the reports (and the benefits for accountability) will continue after programme closure. Whether they do continue is seen as depending on the leadership of senior officers in PO-RALG and the regions, as well as the capacity to aggregate and use this data; until now, it has not been embedded systematically.

WEOs speak very positively about the ways EQUIP-T has changed their work, and as a result they generally want the interventions to continue. WEOs feel that training needs to continue because they would benefit from refresher sessions, and because policies will inevitably change and they need to understand those changes. However, they also feel that even if EQUIP-T ends, they still have the knowledge and skills gained from training so they can continue to perform the job in the same way, and can even train new WEOs if there is turnover.

The extent to which WEOs will continue visiting schools as frequently depends on the resources available, as well as on continued accountability. In some respects, the government took up the idea of WEO grants with the introduction of a responsibility allowance in 2016. However, the EQUIP-T WEO grants were additional, and although some LGAs have included provisions to cover these in their own budgets after EQUIP-T ends, the amounts are small and DEOs are doubtful that the money will come reliably. Two DEOs argued that WEOs will continue visiting schools because they know it is their obligation, and WEOs in other regions have been managing with only the responsibility allowance. However, most WEOs interviewed think they will not have enough money to continue the visits as they have been, and with less funds, WEOs say they will prioritise fuel over maintenance.

For more information, see EL Qual, Sections 6.1 and 6.2.

## 7.4 District component summary

**Table 6. Summary of findings for the district component**

### Was the district component effective at achieving its objectives, and why?

**The component was not effective at strengthening the PFM capacity of district officers, but DEMs have improved management practices in the districts. The component was effective at strengthening WEOs' capacity and increasing the frequency of their visits to schools.**

The LGAs visited were selected because EQUIP-T staff identified them as successful cases, where it could be expected to find the best practices and strongest success stories. The reality in these three LGAs shows mixed achievement of the component 3 interventions on improving district planning and management; given that these were identified as success stories, it means even lower achievement might be expected elsewhere. In these cases, EQUIP-T has been effective at building awareness and understanding of LGA officers on education issues, at introducing DEMs as a method for collaborative and efficient education management. However, the programme has been relatively unsuccessful at building the general capacity of LGA officers in planning and budgeting; this is not a surprise given the programme's shift in focus towards implementation through LGAs rather than capacity building, and in this regard the programme has been able to implement a substantial volume of activities through government systems.

In these LGAs, EQUIP-T has been effective at empowering and building the capacity of WEOs and facilitating them to visit schools more frequently, due to the various training opportunities, including DEMs, and the motorbikes and grants. However, a wider sample of head teachers in the quantitative study were less positive about the support received from WEOs, which may relate to the high WEO turnover experienced in 2017.

### Are factors favouring sustainability in place?

**Enthusiasm for sustaining the successful practices is there at local level, so long as commitment and leadership is shown from higher levels of government.**

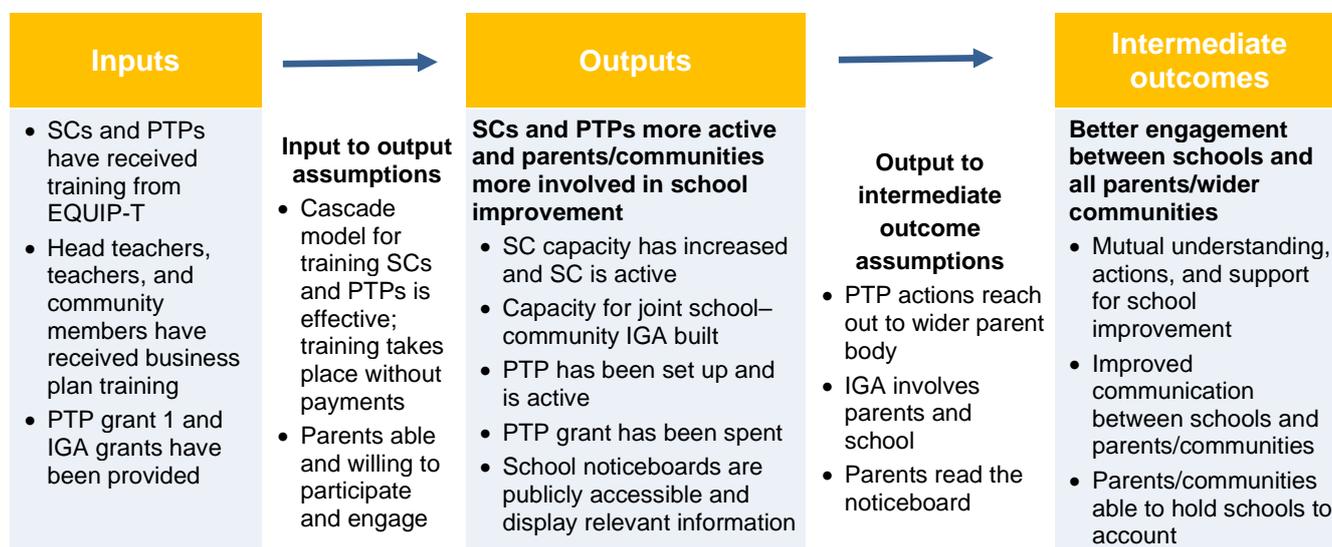
At LGA level, DEMs and WEO training and grants are seen as effective and the capacity is there to continue with them. However, there will need to be commitment and leadership from more senior levels to sustain these activities and the benefits, such as more frequent school visits. The challenge for LGAs in affording additional resources for WEOs will affect the frequency of school visits and the attendance of WEOs at DEMs.

## 8 Community component

This chapter contains the findings on component 4, which has two sub-components: 4A, on community participation and accountability; and 4B, on a conducive learning environment for marginalised children, particularly for girls and children with disabilities. The findings correspond to the criteria of effectiveness and sustainability, and draw on the quantitative endline as well as the qualitative endline, which sampled six schools/communities identified as having more successful engagement of PTPs – a core intervention under 4A. The evaluation results chains for 4A and 4B are described here.

Sub-component 4A aims to build better engagement between schools and parents and the wider community, with the overall outcome objective of strengthened community participation and accountability. The inputs are training for SCs, the establishment of PTPs, the provision of PTP grants<sup>46</sup> and IGA grants,<sup>47</sup> the distribution of school noticeboards, and the carrying out of CENAs.<sup>48</sup> These would lead to outputs: more active and capacitated SCs; established and active PTPs; spending of PTP grants and the establishment of IGAs; information displayed on noticeboards; and a working CENA action plan. Together these are expected to lead to a strengthened relationship of mutual support between schools and parents for school improvement, strengthened accountability, and improved communication. Key assumptions here are that the cascade model for training is effective; participants attend training without payments; parents have the time, interest, and motivation to engage (with PTPs and with the school more generally); and parents read the noticeboard. The assumptions required to strengthen accountability are not elaborated in EQUIP-T documents and are not comprehensively articulated here. More details on the objectives and activities of PTPs are given in Annex C.

**Figure 15. Main results chain for EQUIP-T sub-component 4A community participation and accountability**



<sup>46</sup> Schools received TZS 550,000 as PTP grant 1, of which TZS 100,000 was for PTP activities and the remainder for general school improvement.

<sup>47</sup> EQUIP-T provided IGA grants of TZS 1,500,000 to 50% of schools in each district.

<sup>48</sup> This sub-component also includes distribution of school scorecards which were provided after the quantitative fieldwork.

- School noticeboards have been distributed by EQUIP-T
- CENAs have been conducted

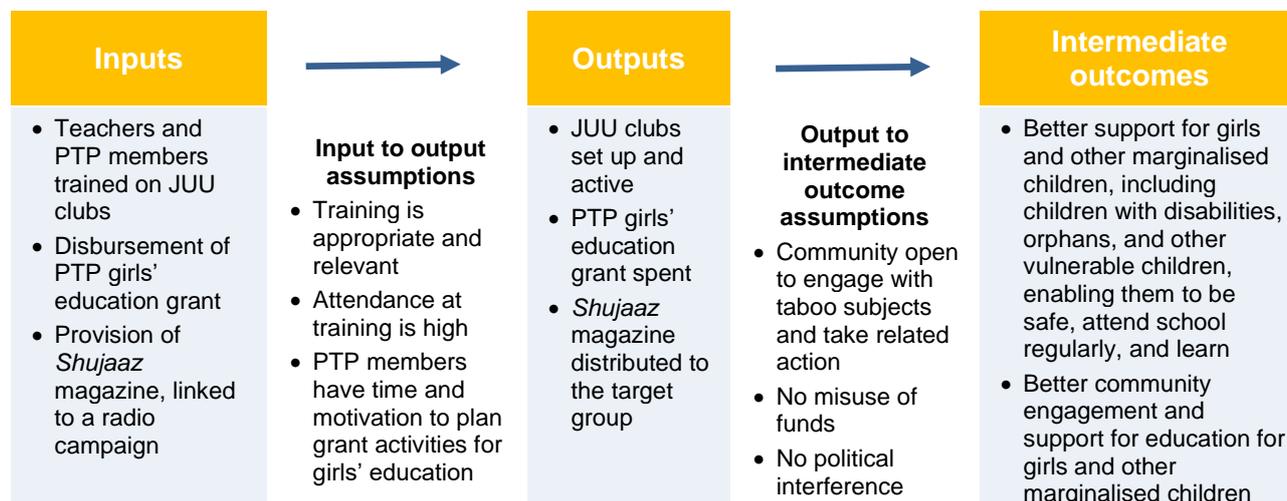
- CENA action plan is in place and is linked to SDP

- Parents able and willing to participate and engage

EQUIP-T’s logframe indicators which relate to this results chain are: Output indicators: (1) Percentage of schools with PTPs participating in activities to improve inclusion (related also to 4B); (2) Percentage of schools with scorecard displaying updated information (this activity began after the quantitative endline survey), updated in 2019 to the percentage of PTPs taking action based on scorecard results; and (3) Percentage of schools continuing IGAs using their IGA grant.

Under sub-component 4B, training was provided for teachers and PTPs on setting up JUU clubs,<sup>49</sup> a second PTP grant was provided to focus on girls’ education,<sup>50</sup> and *Shujaaz* magazine<sup>51</sup> was distributed to schools.<sup>52</sup> As outputs, the component is expected to see active JUU clubs, the PTP grant 2 spent, and magazines distributed. The intermediate outcome objectives are better support from the school and community to help girls and marginalised children be safe, attend school, and learn. Sub-component 4B was not measured at the outcome level, and though not a focus of the qualitative endline, one of its main inputs – the second PTP grant – was included in research into PTP grants.

**Figure 16. Main results chain for EQUIP-T sub-component 4B conducive learning environments**



EQUIP-T’s logframe indicators for this results chain are: Outcome indicator: (1) Percentage of standard VII girl students progressing to secondary school. Output indicators: (1) Percentage of schools with active JUU Clubs, updated in 2019 to percentage of schools implementing JUU clubs pack activities; (2) Number of schools with functional suggestion boxes in which children are raising their concerns; (3) Number of community members reached with social and behavioural change messages on girls education, transition and disability; and (4) Number of districts implementing activities to reduce Female Genital Mutilation (FGM) using grants.

The full evaluation framework is included in Annex B, with component 4 questions in Table 13.

<sup>49</sup> JUU clubs are non-academic and non-extracurricular school clubs for pupils in the upper standards. JUU stands for ‘*Jiamini Uwezo Unao*’ in Kiswahili, which is a motivational slogan that translates as: ‘Be confident, you have the ability, you can do it.’

<sup>50</sup> PTP grant 2, also TZS 550,000, was 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.

<sup>51</sup> *Shujaaz* magazine is a colourful publication that focuses on a young girl and the challenges she faces to attain education.

<sup>52</sup> Later inputs which were not fully implemented at the time of the quantitative survey were suggestion boxes, Social and Behaviour Change Communication (SBCC) messages on girls’ education, transition and disability, and grants aimed at reducing FGM.

## 8.1 Did the community component achieve its outcome targets for community engagement?

**There are still challenges with the relationship between schools and communities, with negative views on both sides that persist due to poor communication. There have been pockets of improvement with PTP and SC members participating in supporting the PTP grants and IGA grant projects, but accountability between schools and communities has not improved.**

There continues to be a view, from head teachers, teachers, SC members, and some community leaders, that parents who do not visit the school, do not do so because they are lazy, uninterested in their children's future, or uneducated. Head teachers' ratings of community support for education is low: only 23% consider it 'good' or 'very good' at endline, and this has not changed significantly over time.

From the parents' perspective, the qualitative endline found a number of occasions where parents did not feel able to communicate with the school or hold the school accountable. Parents did not feel that they could approach teachers to report their concerns, such as about the IGA project, or about dissatisfaction with corporal punishment and fines for pupil absenteeism. At midline, parents expressed doubts about their own ability to raise credible arguments with the school, and felt that the school did not listen to them anyway. These are demonstrative of the challenge of **low self-efficacy**, which has been identified as a barrier to community engagement in improvements in public services in other studies (Lieberman and Zhou, 2019).<sup>53</sup> This is discussed further in Section 8.2.3. Parents see the SC, which has parent membership, as formally responsible for representing their views at the school, and thus do not feel they should engage directly.

On the other hand, parents view the school as belonging to the community and feel that it is the community's collective responsibility to support the school. This corresponds with a national study in which citizens affirm the importance of parental involvement in the education sector.<sup>54</sup> Parents also see themselves as responsible for ensuring that their child is clean, and in school on time, as well as for following up with their children's school and homework – these perceptions could not be attributed to changes made by EQUIP-T. Meanwhile, a sizable share of parents (40%) did not receive any written information or meet a teacher about their child's academic progress during the school year.

**Both the PTP grants and the IGA grants' projects gave the SC and PTP an avenue and purpose to communicate with parents, asking them for their support in school improvement.** Schools and parents have, in many cases, demanded accountability and transparency from each other in relation to these grants.

A positive outcome of PTP grants has been in mobilising the community to provide money and labour to improve school infrastructure. PTP grants have been welcomed as schools have needed more infrastructure to cope with increased enrolment following fee-free education. In some schools, the grant was used for activities to improve learning, such as academic camps, and PTPs played a role in encouraging parents to support this. Participating in PTP activities also improves the PTP members' sense of ownership of, and pride in working with, the school.

<sup>53</sup> Lieberman and Zhou (2019) argue that low levels of citizen engagement in improving service delivery is due to low self-efficacy: a person's sense of being capable of affecting change.

<sup>54</sup> In its Sauti za Wananchi study in 2016, Twaweza found that 85% of citizens cite lack of parental involvement as a serious problem facing the education sector (Twaweza, 2016).

IGA projects were well received by parents and community members and were viewed as benefiting the community and school development, and as providing food programmes for pupils. Some parents became involved in the IGA projects because they believed in their potential to support the school environment and therefore their children's education. Many parents expressed ownership of the IGA project. Parents report being engaged by the SC, PTP, and head teacher for their support through contributions and labour, such as making bricks, helping with construction, and taking care of goats. Parents have also then taken an interest in visiting the school to check the on the IGA's progress.

With the introduction of PTPs, the modest improvement in information sharing between schools and parents has been driven by the search for community support for PTP grant and IGA projects, and the need to raise parental awareness of the importance of education for their children. Parents do not feel supported by the PTP or SC on broader issues, such as sharing dissatisfaction about fines for absent pupils with head teachers, village leaders, or WEOs, nor do PTPs or SCs assuage parents' fears of being fined. In several schools, SCs participated in the decision to levy fines against parents whose children are absent, and SCs and PTPs did not support parents in raising concerns with the school leadership. This indicates a limitation in the ability of the PTP and SC to bridge the communication gap and thus strengthen accountability.

Relating to sub-component 4B, the logframe reports an outcome indicator of the proportion of Standard 7 girl students progressing to secondary school; in 2019, this was 65% in the original seven regions, up from 51% in 2016.<sup>55</sup> Whilst the logframe reports the indicator, it does not include a target to assess this performance against.

For more information, see EL Qual, Sections 4.1, 4.2 and EL Quant Vol 1, Section 6.7, as well as Annex C.

## 8.2 What factors explain the results?

### 8.2.1 Were inputs (training on SCs, PTPs, and grants; PTP and IGA grants; noticeboards, CENAs, *Shujaz* magazine, and JUU club) delivered as intended?

**Some inputs for the community engagement sub-component were delivered as intended; however, the training for PTPs to carry out their responsibilities was not implemented to plan. There was also variation in how training for the SC was cascaded within the school, and incomplete coverage of CENAs.<sup>56</sup>**

The training of SCs and PTPs, including on spending the PTP and IGA grants, relies on a cascade model in which head teachers and WEOs are trained first by LGA officers and then pass on what they have learned to SC members and PTP members. At endline, 76% of head teachers had received training on SC responsibilities and roles; 60% on PTP responsibilities; 60% on the application and management of PTP grant 1 and 41% on grant 2; and 56% on how to develop a business plan for income generation. In terms of cascading down, significantly more SCs (87%) had received some training in the previous two years compared to midline (72%). Training sessions across the qualitative sampled schools were run by WEOs, with the head teacher also in attendance, and in almost all schools training was a one-off event. For PTPs, in-school training for PTP parent and teacher members also increased significantly, with 72% of PTPs having received training at endline, compared with 39% at midline. However, this training was not implemented consistently, both in terms of the

<sup>55</sup> It is not clear which regions the 2016 data relates to.

<sup>56</sup> The evaluation was conducted before implementation of the community scorecards had begun.

duration of the training and the content covered. Head teachers reported the short duration of training to be a challenge, given the amount of content covered, but overall were happy with PTP training provided to them. Similarly, PTP and SC members felt there was insufficient time in training to cover some of the skills needed, especially in relation to supervising and handling school finances. The main reason for not conducting training or holding training for shorter periods in school for PTP and SC members was the lack of incentives for participants to visit the school and participate for the entire training duration.

At endline, all schools had received PTP grant 1, and virtually all of them received the correct amount.

The EQUIP-T target of providing IGA grants to 50% of all schools in the programme districts had essentially been achieved at endline. The IGA business plans were accepted for 49% of the programme schools surveyed, and close to all schools (98%) whose business plans were accepted received the full amount of the IGA grant. Before this, almost all school-community groups (96%) had developed and submitted a business plan for IGAs, suggesting that at least 96% of school-community groups had received training on the business plans.

Based on head teacher reporting, 88% of schools received a noticeboard from EQUIP-T in 2016 or 2017, which indicates a sizeable shortfall.

CENAs were conducted in 58% of the school communities during 2014–2017, leaving a substantial proportion of schools which did not receive this input.<sup>57</sup>

**The various interventions under sub-component 4B related to the JUU clubs, PTP girls' education grants, and *Shujaaz* magazines have been delivered to schools largely as intended.** Close to or over 90% of schools report receiving training in 2016 or 2017 on setting up and running a JUU club, receiving the PTP girls' education grant in full, and receiving copies of *Shujaaz* magazine in 2016 or 2017.

For more information, see EL Qual, Sections 4.1 and 4.3, and EL Quant Vol. 1, Sections 5.3, 6.2, 6.4, 6.5, 6.6, and 7.3.

### 8.2.2 Were outputs of the community component achieved as intended? If not, why not?

**SCs in the qualitative endline report that they have a greater understanding of their roles, and are more motivated, as a result of the EQUIP-T training.** SC members reported that the training they received was beneficial to understanding the importance of their roles in schools, and this information made SCs aware of responsibilities they were not yet fulfilling. Although SCs are not meeting more regularly (69% met during the last quarter at the quantitative endline – not a significant change from midline), SCs in the qualitative endline became more active in supervision of the school, the budget, and school development, and became more active in communicating with parents about education. At both midline and endline, qualitative interviews suggest that SC performance was seen as improving over the last two to five years, but the broader quantitative survey found no change in head teacher rating of SC contributions and support since midline and baseline.

<sup>57</sup> The EQUIP-T annual report for 2017 states that CENA had been rolled out first to rural and then to municipal districts in the original five regions. The low proportion reported in the quantitative survey could indicate that schools and head teachers were not involved in the CENA, or that (due to turnover) the head teacher was not aware of the CENA, suggesting it had not had lasting effects.

Overall, almost all parents and community leaders know the existence of the SC and can identify outcomes of the committee's work in detail. Many parents perceive the SC as serving their interests since some members are chosen and elected by parents. Parents view the SC as an avenue to address challenges in the school, following up on teachers' progress and issues such as excessive punishment of pupils.

**PTPs have been set up broadly as intended, but the process of selecting PTP parent members was not always in line with the plan.** At quantitative endline, virtually all schools (99%) have a PTP, a significant change from baseline, when only 14% of schools had a parent–teacher group. PTPs are formed as intended, with on average seven fathers, seven mothers, and seven teachers. The qualitative endline found that the teachers, head teacher and SC often drive the selection process by shortlisting the PTP members first, whereas the intention was for parents to put themselves forward first. Head teachers, teachers, and SCs preferred to choose parents who were already influential and active in the school and other community activities; these individuals tend to be economically better off and live closer to the school, so creating a bias towards which parents are selected onto the PTP.

**It is hard to reach a conclusion on the success of PTPs at achieving the intended outputs, given the broad definition of PTPs' objectives in programme documentation and lack of detail on the expected mechanisms between activities and results.** Clearly, PTPs are expected to be active in some way, and in this respect there has been some success: the proportion of PTPs which took some action to improve education increased significantly, from 47% of schools at midline to 68% at endline, though this means almost one third were still not active. The midline study found that the engagement of the head teacher in a supervisory role was associated with a more active PTP.

The intended activities of PTPs are not predetermined – they are 'meant to be decided at school level based on each school's needs and priorities' (EQUIP-T, no date, p. 2);<sup>58</sup> however, programme documentation does include a number of suggested activities (see Annex C for a detailed review). Of these, the category of activities most commonly carried out by PTPs is action to follow up and reduce pupil absenteeism. Nearly half of schools' PTPs took this type of action; examples from successful PTPs included checking attendance in classrooms, visiting, or writing to parents of absent pupils, speaking to community leaders, and taking action to directly bring truant children back to school. Around one in 10 schools' PTPs took action around infrastructure (though this was often linked to use of the PTP grants), and organising school catering. In schools identified as having successful PTPs, PTPs had played a role in sensitising girls and parents on the use of sanitary pads (often linked to purchasing pads with the second PTP grant) and on avoiding early pregnancy. Whilst PTPs do not appear to hold fundraising events, the successful PTPs have often been involved in collecting contributions from parents (such as for school feeding or construction). Across this range of activities, the interaction with individual parents and community leaders suggests there are actions to increase information flow between the school and wider parents and community (beyond the 14 parent members of the PTP), usually around advocating more support for their children's education. It is relatively uncommon for PTPs to help directly in the classroom, but successful PTPs report that they help with discipline, check whether pupils are learning, discuss pupil progress with teachers, and take a role in finding solutions to the problem of teacher shortages.

The frequency of meetings – another measure of activeness of PTPs – is low. On average, to head teachers' knowledge, schools held fewer than two PTP meetings in 2017. However, it is possible that more frequent informal meetings took place between sub-sets of PTP members. In some schools, meetings took place between the head teacher and PTP teachers without parents, missing an

<sup>58</sup> This is the PTP Practice Paper, produced in 2019 or 2020.

opportunity for the school to engage parents systematically in the school's affairs. At the same time, parent members sometimes meet without including the teachers.

The programme logframe has two 2019 output targets relating to PTPs' activities. One is the proportion of PTPs taking actions based on scorecard results (scorecards were distributed after the quantitative fieldwork). The target for 2019 is 75% in the original seven regions, and the result is reported as 76%. The logframe proposes a reduction in targets for the new regions due to delays with tranche disbursement, which have shortened the time for implementation. This evaluation has not collected data on the school report cards given that these were implemented after the quantitative endline. The other indicator is the proportion of schools with PTPs participating in activities to improve inclusion. Performance, as reported in the logframe, is far behind target in 2019, with only 37% achieved in the original seven regions and 44% in the two new regions, against targets of 70% and 50%, respectively.

**PTP grants were spent in nearly all schools by endline, but PTP parent members were not as closely involved in planning for the PTP grants as they could have been.** Most head teachers in the qualitative endline decided how to spend the grants with the teachers of the school, involving the SC for approval. As grants came with guidelines, only a few head teachers worked collectively with the PTP parent members to decide how to spend the grant. As the SC has responsibility for school finances, some SCs asked PTPs to rubber-stamp the PTP grant plans. It is discouraging to find that decisions for improvement are still primarily led by the school. If PTP members do not feel empowered to participate in spending grants or providing feedback to the head teacher and SC, it is unlikely that they will be able to empower or engage other parents to do so. The most common usage of PTP grant 1, which was fully spent by endline, is for school infrastructure and furniture (93% of schools) and extra-curricular activities (61%), in line with the allowable uses of these grants.

At the time of the quantitative endline survey, only a quarter of schools had spent PTP grant 2, so the sample is too small to understand what schools chose to spend the grant on. The PTP girls' education grant was intended 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. Of all schools which had spent the grant, 66% had spent some of it on pupil welfare activities, 66% spent some on infrastructure and furniture, and 35% spent it on extra-curricular activities. Additionally, 70% of schools which had spent the grant state that the expenditures were specifically to support girls, whilst the remaining 30% report that the grant expenditures were not specifically for any group of vulnerable pupils. In most schools visited during the qualitative endline, the PTP girls' education grant was used to buy sanitary napkins or soap for girls, or for building a changing room for them to provide them with some privacy.

**In schools with IGA grants, SCs describe being actively involved in the conceptualisation, decision making, and implementation of the IGA.** Among the schools that received IGA grants, 88% had started at least some IGAs by the quantitative endline.

The logframe includes an output indicator for the 'proportion of schools continuing with Income Generating Activities using their IGA Grant'. The logframe reports that this was almost met, with 71% of schools that received a grant continuing activities in 2019, against a target of 75% (in the seven original regions).

**After an initial significant increase, the proportion of schools with noticeboards that are publicly displayed returned to its initial share by endline.** More than half of the displayed noticeboards (49% are displayed) contain information about academic results, and teaching and learning; a third contain SDP, budget, grants, and other financial information. One possible reason for the decline in displayed noticeboards between midline and endline might be that the noticeboards

provided were not particularly suitable for display outdoors as their surface was made of fabric, without protective covers or a shelter. The noticeboards were in line with Government of Tanzania standards, but this limitation would have been apparent to the programme when procuring them. Alternatively, head teachers may have found it inconvenient to move them in and out of a locked room each day or be concerned about them being stolen. The midline qualitative fieldwork found an indication that the noticeboards were displayed just in time for the research team's visit.

It appears that when communities undertake a CENA, a large proportion (77%) of them take some action based on it. The most common actions were to improve school infrastructure.

**JUU clubs have been established in the large majority of schools (91%), and the majority of these (79%) are active.** The composition of the JUU clubs are, on average, gender-balanced; they are typically made up of about 16 male pupils and 17 female pupils. On average, JUU clubs met five times in 2017. The activities largely related to the school and learning environment; fewer were targeted at addressing issues of health and hygiene of pupils or their right to education. A minority of JUU clubs have targeted groups of vulnerable pupils in their activities, which would be expected to be a priority for the provision of better support to girls and other disadvantaged children. In 2019, the programme was behind on its logframe target for the 'proportion of schools implementing JUU clubs pack activities'; there was zero actual achievement against targets of 75% and 60% in the seven initial and two new regions.

One logframe indicator for this component is the number of schools with functional suggestion boxes for children to raise their concerns, measured by the WEO-administered survey. For this, the achievement is 53 schools, against a target of 120 schools – so far behind target. In the evaluation's quantitative endline, a suggestion box was 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.

The logframe has two output indicators for sub-component 4B which were not covered by this evaluation. The first is the number of community members reached with SBCC messages on girls' education, transition, and disability. Here the total target is 14,000 but no community members had been reached in 2019. The second is the number of districts implementing activities to reduce FGM using grants – again, the achievement is zero, against a target of 20 districts. The poor performance of these indicators is related to delays in implementation – the quarterly report for April to June 2019 (EQUIP-T MA, 2019) states that the SBCC interventions were still being designed, and the FGM grants (LGA grants to carry out activities to stop FGM) were part of the seventh tranche of LGA funds, which was still delayed in June 2019.

For more information, see EL Qual, Sections 4.1 and 4.2, and EL Quant Vol 1, Sections 6.2, 6.3, 6.4, 6.5, 6.6, and 7.3, as well as Annex C.

### 8.2.3 Did the assumptions required for outputs to lead to outcomes for the community component hold?

**Schools organise activities involving SCs, PTP parent members, or the wider group of parents, according to perceptions of their willingness to take part. This approach creates actual barriers for participants.** Head teachers believe parents would not be willing to visit the school often, due to lack of allowances, the distance, and the need to miss work to attend these meetings. Head teachers and WEOs also believe that teachers have a better understanding of the school's needs than parents. These views are corroborated by members/parents: they expect explicit communication about compensation for attending meetings, and sometimes will not attend due to the opportunity cost of

foregone income. Distance and topography create difficulties, and the division between active and inactive members is sometimes along geographical lines.

These issues have manifested themselves in a number of ways. After their initial training, head teachers went on to provide less than comprehensive training to PTP members, and may not have consulted PTPs in PTP grant planning. Schools do not hold frequent PTP meetings for the same reasons (lack of funds for allowances, concerns about participation), and attendance at SC meetings is affected by similar issues. PTP members are also unclear about who would train new members, posing a risk to the sustained activity of the PTP.

**For the schools that do have active PTPs, the success of these activities in leading to the intended outcomes is not clear.** Given that the ultimate objective is to increase community participation and accountability, it is expected that the core PTP membership (seven teachers and 14 parents) reaches out to the wider parental (and community) body and acts as an intermediary for information exchange. However, PTPs have not established themselves as the main link between teachers and other parents, given that at endline only 23% of parents (of Standard 3 pupils) know that the PTP exists. Furthermore, parents who live far from the school or visit infrequently are less likely to know of it, and those who do know are unclear what its role is. The reasons for the PTPs' relative obscurity could be lack of action (with around one third of PTPs inactive), the failure of their actions to engage parents more widely, or parents not knowing that these actions are the result of the PTP. On the other hand, if the objective is simply that, by means of these 14 parent members, the PTP brings some parents closer to the classroom, then there has been moderate success, with two thirds of PTPs being active in at least some way.

The conditions needed to lead to greater accountability are not clearly defined and may also not have been met. Increased flow of information – as intended by the actions of the PTP – may be a necessary condition but not sufficient one, as evidenced by studies aiming to increase citizen engagement via greater information and transparency (Lieberman and Zhou, 2019). Whilst PTPs may be active in some senses, the conditions needed to empower even PTP parent members (let alone the wider community) to challenge and hold the school to account are not there – and have not been changed by PTPs. The programme design does not clearly articulate what mechanisms are expected to lead from the PTP activities to the intended outcome (see Annex C).

One such bottleneck to greater accountability may be low self-efficacy: a person's sense of being capable of affecting change (Lieberman and Zhou, 2019). This relates both to internal efficacy – the belief that one can participate – and external efficacy – the belief that authorities, in this case teachers and education leaders, will respond. The midline research found examples that relate to this. In terms of internal efficacy, parents explained they felt unable to comment on academic matters since they were uneducated themselves. In this respect, teachers are seen as superior, and parents felt that teachers look down on them. Parents felt that schools do not listen to their opinions, and consult parents only when they know there will be agreement – a sign of low external efficacy. If parents' efforts to participate are not affirmed through a positive response by authorities (teachers and education leaders), they may fall into a trap of low self-efficacy and thus believe it is not worth participating. Finally, at midline, parents spoke of their fear of experiencing repercussions from holding teachers to account. This could either lead to their child being excluded, or the teacher eventually requesting a transfer – a serious threat for schools that struggle to retain enough teachers.

**Those PTPs that are active tend to be focused on improving pupil attendance and punctuality.**

In the qualitative sample, PTPs believe they play an important role in improving communication with parents whose children have been absent. The head teachers appreciate the role PTP parent members play in increasing awareness of the importance of schooling and visiting parents whose children have been absent to convince them to come to school. Conversely, whilst some parents know

the PTP as the people who follow up on absenteeism, on the whole they do not corroborate the view that the PTP has supported schools in increasing pupil attendance. SCs are also seen as having a role in reducing pupil absenteeism. Despite a large proportion of PTPs acting 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.

**Whilst – in the most positive cases – there may not be tension between the SC and PTP, the SC is recognised as having more authority and clearer roles.** In schools sampled because they were success cases for PTP activity, the PTP and SC report their relationship to be mutually supportive and complementary in addressing critical challenges in the school. However, SCs are generally viewed as being the ‘owner of the school’, a body authorised by the government with a long history in the school, and the PTP is the ‘sub-committee’, which reduces the likelihood that PTPs will be empowered to represent parents in holding the school to account. The SC has more authority in the school than the PTP or parents to ensure teachers are fulfilling their duties and in conducting checks on teaching. PTP members accept their role as being subordinate but complementary to the SC, and there was no tension observed during the qualitative endline. More generally, there is some confusion about the different roles of SCs and PTPs, with 83% of head teachers seeing them as similar or only somewhat different in the quantitative endline, a view that was corroborated by head teachers as well as by some SC and PTP members in the qualitative endline. The PTP’s position is less widely recognised than the SC’s: it is not clear from the design what the intended authority of PTPs was (see Annex C), and PTP members themselves do not believe they have authority. Thus, the additional and distinct value that PTPs could play in improving wider parent–school relationships and accountability is not evident.

**The school noticeboard has made only a small contribution towards improving information sharing between schools and parents.** Only 29% of parents are aware that the school has a noticeboard, and a mere 18% of parents had read it in the last quarter. Parents who are illiterate are significantly less likely to know about the noticeboard, though distance from school does not appear to affect the likelihood of knowing about the noticeboard.

For more information, see EL Qual, Sections 4.1 and 4.2, EL Quant Vol. 1, Section 6.3 and 6.5, and the Midline IE report Section 6.2.

#### 8.2.4 What unanticipated and unavoidable factors affected the results?

As with other components, the announcement of the fee-free education policy had some effect on this component. The expansion in enrolment meant there was a greater need for school infrastructure, which reinforced the need for community contributions of labour or in-kind resources.

### 8.3 To what extent are factors favouring sustainability in place?

#### 8.3.1 Was the component considered effective? Are key stakeholders committed to continuing these activities and practices? Is the capacity in place to continue the activities?

At the school and community level, even in schools identified as having more active PTPs, the component is not effective at improving school and community relationships. The lack of trust and understanding between schools and parents and the lack of an effective communication platform even at endline suggest that there has been little benefit of this component at scale. Therefore, whilst it is important to bridge the communication gap between schools and communities, there is little benefit in continuing this component in its current design.

The component is effective in directing funds and efforts towards school infrastructure improvement, and to a lesser degree, improving access to sanitary products and facilities for girls. PTP members report that they want to continue supporting school development, but that funding from the PTP and IGA grants is necessary for this. Any profits from IGAs could fund continued activity. However, PTPs and SCs feel they need support from experts, such as the individuals involved with EQUIP-T, to guide them, suggesting they may not have the commitment and capacity to continue on their own.

Continuation of active or more effective PTPs and SCs will also depend on training for new members given the turnover that takes place. If WEOs and head teachers provided the initial training, it should be possible for them to provide further rounds of training (subject to their own turnover); however, this is not happening in practice. There is no clarity about the responsibility to continue holding training, and WEOs say they will but are concerned about their ability to do so in the future without resources. Furthermore, they have not been active in running repeat training in the past.

At national level, the government has selected PTPs as one of its best practices, along with IGAs, as a revival of education for self-reliance.<sup>59</sup> In a letter to EQUIP-T in January 2019, PO-RALG committed to continuing and scaling-up PTPs and IGAs, along with four other activities. However, the government does not expect to provide the grants, which appear to be a precondition for mobilising community support and activity; therefore, this activity is unlikely to be sustained in the future.

For more information, see EL Qual, Section 4.4.

### 8.3.2 Is it affordable for the government to continue to replicate and scale up PTPs and IGAs?

**Continuing the establishment of PTPs and scaling-up PTP and IGA grants in all regions is relatively affordable. These two interventions would cost less than 3% of the annual LGA non-salary budget.**

The endline cost study estimated the costs of continuing with PTPs and IGA in the existing regions and scaling-up to the other regions of the country, over a five-year period.<sup>60</sup> The analysis included a maximum scenario, which for both PTPs and IGA involves including the PTP grants and IGA grants for schools (once) in new regions where it is scaling up. The grants are included in a maximum scenario to reflect the cost of scaling-up these activities in a model close to that used by EQUIP-T. This is particularly important given the main benefit of this component has come about as a result of having the PTP grants and IGA grants to drive community mobilisation and communication efforts. The maximum scenario also includes a refresher training in all regions on the set-up and roles of PTPs, and on IGA projects, every three years. Reflecting that the government does not intend to distribute grants, the minimum scenario does not include grants or refresher training, and thus the main cost results from rolling-out initial training in the first year in new regions.

**For both PTPs and IGA, the cost of the grants themselves is a relatively small portion of the cost of these interventions. The larger cost comes from holding training sessions on the activities, and the opportunity cost of government staff time spent on these activities.**

<sup>59</sup> Education for self-reliance was a concept introduced by Julius Nyerere, the first President of Tanzania, in 1967. It involves the idea of giving skills and knowledge children to allow them to be independent in their futures. IGA projects are one way for pupils to learn life (and income-generating) skills.

<sup>60</sup> In the case of PTPs, the existing regions are all nine EQUIP-T regions, plus four regions which received PTPs under the Tusome Pamoja programme. For IGA, it is only the initial seven regions.

For PTPs, distributing PTP grants to schools in new regions in the first year of scale-up would cost over TZS 9 billion (approximately £3 million), and whilst the initial and refresher training sessions incur costs, in many years there is very little cost. The average annual cost over five years would be TZS 6.7 billion, which includes the costs of salaries of government staff preparing for and attending training sessions and parent meetings. If the salary cost is assumed to be absorbed within existing salaries, then the average annual additional cost would be TZS 3.4 billion, of which TZS 2 billion relates to the grant payments. In the minimum scenario, with no grants and no refresher training, the average cost would be TZS 3.3 billion per year, only TZS 370 million of which needs to be budgeted if salaries are excluded.

For IGA, as with PTPs, the initial distribution of grants to 50% of schools in new regions costs TZS 9 billion. With training and grant application, the total cost of the scale-up in the first year would be TZS 28 billion. The average annual cost of IGA in the maximum scenario is nearly TZS 11 billion. If salaries are excluded, the government would need to budget for TZS 8 billion on average each year. In the minimum scenario, the cost is only a one-off initial roll-out of training which costs TZS 11 billion, or less than TZS 3 billion per year annually. However, this scenario may not secure the success of community relationships or the necessary seed money to run IGA projects without the grants.

Put together, the maximum scenario for PTPs and IGAs would cost TZS 17.5 billion each year on average, or TZS 11 billion without salaries. The minimum scenario would cost TZS 6 billion per year on average, or TZS 2.5 billion if salaries are excluded.

To put this in context, in 2018/19 the total non-salary budget for education in the LGAs was TZS 422 billion. Assuming the salary costs are absorbed by existing staff time, only the non-salary costs are additional. In the maximum scenario, the annual average non-salary costs of PTPs and IGA are 2.7% of total LGA non-salary budget, and in the minimum scenario, 0.6%.

As discussed in Section 5.3.2 regarding the costs of the in-service training model, LGAs have very little room for discretion in spending their non-salary budgets. The costs of PTPs and IGAs is much lower than that of in-service training, and so more possible to afford, but will still be a challenge given the various priorities competing for resources. It is unlikely that LGAs would choose to make room to fund these activities within the current budget they receive, without substantial directives from above. As with the funding of the in-service training, the government could look to increase total allocation to the education sector in order to cover these activities. The costs are relatively very small compared with the total sector and the potential increase in government revenue. However, whilst this evaluation has found the PTP grant and IGA grant to be useful for schools, the establishment of the PTP itself is not valuable, and is not a requirement for scaling up some form of school grant.

For more information, see EL Cost Sections 4 and 5.

## 8.4 Community component summary

**Table 7. Summary of findings for the community component**

**Was the community component effective at achieving its objectives, and why?**

**No, the community component has not been effective in achieving its objective of improving community participation and accountability with schools.**

There have been pockets of improvement in the relationship, support, and communication between schools and communities; however, at an aggregate level, there are still serious challenges and the general situation has not improved substantially.

The PTP grants and IGA grant projects provided a trigger for more – although still limited – communication between the school and parents, often channelled through the SC and PTP. In successful cases, parents and the community did mobilise to support these projects with contributions, labour, and a sense of ownership. However, PTPs have not added more generally to an improvement in communication, relationships, and accountability, as seen by the fact that most parents do not know of the PTPs' existence. The reasons for this are many: parents are unwilling to give up their time, given the opportunity cost of doing so, and the PTP is seen as subordinate to the SC and therefore as having less authority and as less empowered to act. There is dissatisfaction against fines levied on parents of absent students and corporal punishment, which have not been addressed. The assumptions required for the interventions to lead to greater accountability were not well defined in the programme design, and barriers include parents' lack of self-belief and the view that schools would not take action anyway. At a lower level of ambition, parent members of PTPs in around two thirds of schools have been more actively engaged in school matters.

On the other hand, the training for SC members did improve their understanding and motivation for carrying out their roles. SCs appear to have been more actively involved in the decisions around IGA grants and PTP grants than SC members – reflecting their mandate for overseeing school finances – and are known by the community as being responsible for representing parents' views at the school. This thus raises the question of whether the SC could fulfil (at least some of) the intended aims of the PTP more efficiently and with less confusion.

The school noticeboards have made only a small contribution to improving information flow, given that only half are displayed, and most parents do not know they exist.

Ultimately, some of the underlying challenges – teachers' negative views of parents, who are seen as lazy and uninterested; and parents feeling unable to approach teachers or hold the school to account – are substantial and will take a long time to change.

The majority of JUU clubs are active, but this evaluation has not been able to assess how effective they have been at improving support for girls and other marginalised children.

### Are factors favouring sustainability in place?

**With little benefit felt at school level, the benefits and drive to continue the activities will not be sustained. Leadership from senior government would be needed to continue the activities, including some budget for distributing grants, which are the catalyst for community mobilisation.**

At the school level, the benefits from this component are not substantial and thus the motivation for the school and community to continue with the activities (in particular, setting up and running a PTP) are low. That said, PTP members report that they want to continue in their role, but the funding from PTP and IGA grants is critical for sustaining this enthusiasm. New PTP and SC members will need to receive training and schools are not clear who has responsibility for this, meaning there is a risk that it will not continue unless directives come from senior government leadership through the LGA level.

The government has committed to continuing and scaling-up PTPs and IGAs as best practice activities. Doing this, with grants in new regions and refresher training, would cost 2.7% of LGAs' non-salary budget, which may be a challenge unless additional budget is provided with directives to use the money for these activities. If there are no grants or refresher training, the cost is very low as it is just an initial training roll-out. However, the efficacy of the spending is questionable, given that the grants appear to be a precondition for mobilising community support and activity.

## 9 Conclusions, recommendations, and lessons

This chapter draws together the key conclusions from this evaluation which are most relevant for the Government of Tanzania and DFID. However, they will also be of interest to other education stakeholders in Tanzania who are developing, supporting, and providing education services, and will have relevance for stakeholders in other countries facing similar contexts. The conclusions are followed by recommendations addressed directly to the government and DFID, and lessons which are important messages for development programming and evaluation elsewhere.

### 9.1 Conclusions

The conclusions are structured around the four DAC criteria for which this evaluation has the most evidence: impact; effectiveness; efficiency; and sustainability. The components are each discussed with relation to effectiveness, though of course sustainability relates to all components, sometimes in different ways. At the end of each sub-section, a reference to the relevant findings chapters, underlying the conclusions, is given in parentheses.

#### Impact

**EQUIP-T had a substantial positive impact on learning outcomes, contributing to the programme's overall objective.** In both Kiswahili and maths, the programme helped children in Standard 3 move up from low-performing bands closer towards meeting Standard 2-level competency. The impact was particularly notable for children at the bottom of the performance distribution, helping many more children move up from this very low base than would have been the case in the absence of EQUIP-T. Whilst the logframe targets were not quite achieved, the improved outcomes of the lowest performing pupils are significant and not captured in the logframe.

**This success is all the more notable given some of the contextual factors and changes taking place.** In particular, there were substantial levels of teacher and head teacher turnover, and growing class sizes, each of which were affected by new government policies during the period of EQUIP-T and would put downward pressure on teaching and learning. On the other hand, the change in curriculum in 2015 was complementary to EQUIP-T's focus on numeracy and literacy, and so supported EQUIP-T's aims. Meanwhile, the programme was implemented in regions with a difficult starting environment: a large share of households below the poverty line, low school quality, and large class sizes.

**Despite the large improvement in learning outcomes, it is critical to note that pupil learning outcomes are still very poor, lagging well behind curriculum expectations.** Furthermore, certain groups are particularly disadvantaged: those who do not speak Kiswahili as their home language, and those from poorer households. **(4.2 to 4.4)**

#### Effectiveness

**The evaluation concludes that the main contributor to this impact was the teacher in-service training,** given it had the closest direct link to what children learn in the classroom and was implemented largely as planned. The other components were, to varying degrees, supporting factors, such as through some aspects of better district education management and school improvement activities initiated by grants, but none of the intermediate outcomes have seen unequivocal improvements. **(4.4)**

## Teacher performance

**The teacher in-service training was clearly needed and has been effective.** Although improvement over a range of measures of effective teaching practice has been mixed, the in-service training has helped teachers improve their classroom management, given them a range of tools for use in the classroom, led to more use of teaching aids, and built teacher confidence. The implementation of the school-level training model has faced some challenges which could have been foreseen at design stage, such as complaints about the lack of allowances and fitting the sessions into the school day, yet despite this, the sessions are taking place (albeit with wide variation). For the school-based model to be sustainable, there will need to be commitment and support from all levels of government, and some kind of incentive for teachers to take part, such as building this into the career progression model. It should be noted that the distribution of EQUIP-T's TLMs has not led to effective change, as schools are rarely using them.

As mentioned, staff turnover – from teachers through to LGA staff – is very high and a major issue for the effectiveness of interventions which rely on capacity building and improved relationships. Learning and skills may be lost and the COL ethos may wane as original groups of colleagues disperse. Even at higher levels of government, the turnover in central ministries and regions affects the level of buy-in and thus leadership and commitment to sustaining EQUIP-T activities. A further underlying issue is the severe level of teacher absenteeism from classrooms and the effect this has on teacher and pupil time on task, which is a major factor for children's learning. Whilst time on task was originally in EQUIP-T's logframe, there was never an explicit intervention to address this, and although improvements were seen, it continued to be a problem up to the end of the programme. High teacher turnover and limited time on task – both risks that could have been foreseen in the design phase and depend on wider systemic change – affect the delivery of the intended benefits of the in-service training, and also the sustainability of the training going forward. **(5.1 to 5.3)**

## SLM

**At the time of this evaluation's data collection, the SIS was not being used and thus would need to be assessed and improved in order to make it an effective tool for the use of data in school management** – it is possible some changes have already taken place since 2018. The reasons for lack of use were largely design issues – difficulty and time taken to enter data, and the tablets not working properly. However, poor internet connectivity is also a systemic issue which affects the effective use of the SIS. The other major challenge is that head teachers need sufficient training to use the system, and with the level of turnover, over half of head teachers had not received the training.

Changes in measures of school leadership were mixed, with schools holding more staff meetings and taking actions to address pupil attendance, but whilst head teachers are monitoring lesson plans, they are observing fewer lessons. Only half of schools have an SDP, but most of those that do have started implementing it. The relatively low outcomes from this component are likely due to the high turnover of head teachers (intensified by the new qualification policy in 2017) and head teacher absenteeism, but it also raises the question of the effectiveness of the training content. **(6.1 to 6.2)**

## District management

**In terms of district management, the programme has had little effect in the district office on education planning capacity and processes.** This is not a surprise given the programme's shift in focus towards supporting implementation through LGAs rather than capacity building, and in this regard the programme was able to implement a substantial volume of activities through government systems. There was more positive change in building awareness and understanding of education issues among LGA officers, and in introducing DEMs as a method for collaborative education

management and capacity-building. DEMs are valued by LGA officers and WEOs and are a low-cost activity which could be continued, if encouraged. **In LGAs identified as successful, EQUIP-T has empowered and built the capacity of WEOs** through various training opportunities, including DEMs. It has also facilitated WEOs to visit schools more frequently due to the motorbikes and grants. WEOs' hard work and commitment has also been supported by the government's focus on work ethic, the provision of responsibility allowances, and the higher minimum qualification policy, all of which have elevated WEOs' position and morale. **(7.1 to 7.2)**

### Community participation

**The community component has contributed to pockets of improvement in the relationship, support, and communication between schools and communities. However, at an aggregate level, there are still challenges, and the general situation and levels of accountability have not improved substantially.** Ultimately, some of the underlying challenges – teachers' negative views of parents, who are seen as lazy and uninterested in their children's education; and parents feeling unable to approach teachers or hold the school to account – are substantial and will take a long time to change.

**PTPs have not been successful at bringing the wider parent body closer to schools – they are not well known by parents who are not PTP members.** The space for engagement between schools and communities is led by SCs, which are legally established, but were not ready to share that space, and PTPs were not empowered to play their role in supporting decision making and school improvement. However, some schools were successful in bringing the core parent members of the PTP closer to the school and supporting student achievement. Schools and communities are confused about the difference between PTPs and SCs, and in fact SCs became more active when they received training. Repeating the SC training will be important if their activity is to be sustained. The PTP and IGA grants were instrumental in mobilising parents and community members, including engaging the PTP itself, and it is unlikely the PTPs will survive without the presence of these grants as catalysts. More generally, the issue of allowances or incentives for attending some school activities is likely to become a sticking point for any attempts to raise community involvement. Finally, the school noticeboards have not been effective at improving communication between schools and parents. **(8.1 to 8.2)**

### Design and implementation

**Some of the challenges that have impeded the programme's effectiveness are because of the underlying assumptions in the design; these should have been foreseen.** One example is the difficulty of holding meetings (with teachers or community members) without incentives and fitting these into other schedules; another is the loss of capacity due to high turnover. The challenge of changing cultural norms and behaviour could have been better recognised, particularly in the design of the community component. There are other examples across the components. Although not well documented, it is possible that these challenges were foreseen and considered tolerable. At the same time, the constraints on learning initially identified by the MA are substantial and EQUIP-T's interventions might be expected to make only incremental improvements given the complexities of the system and the length of time needed to effect change. **(4.1, 5.2, 6.2, 8.1, 8.2)**

**However, other reasons for low effectiveness were caused by implementation failings.** Over time, the programme introduced many new activities, sometimes at the expense of others. For example, the shift to decentralised implementation led to less focus on PFM capacity building, and the introduction of the infrastructure sub-component required new expertise in the MA and monitoring systems in LGAs. It is possible that some of the later delays, such as the delayed sub-component 4B activities on inclusion and gender which were part of the original design, were a result of trying to do

too many things.<sup>61</sup> Arguably, the MA spread itself too thin, without an evidence-based analysis of trade-offs. (4.7, 7.1, 7.2, 8.2)

## Efficiency

**The programme's lack of a detailed TOC was a missed opportunity for evidence-based design and reflection.** A detailed TOC would have allowed the MA to precisely recount each of the steps and assumptions, consider the strength of the underlying evidence, and weigh up the risks to success based on the design. There was also no regular critical reflection on a TOC to guide programme adjustments. A more thorough commitment to this exercise (up front and repeatedly) could have led the programme to rule out some of its interventions which were never as likely or did not prove to be effective (such as community components), allowing the reallocation of resources to more effective interventions. The programme was designed to address clear needs in these regions and did adapt to changes in context, as seen, for example, by the addition of an infrastructure component in 2017. Whilst the programme was intended to provide broad systemic support, rather than being a narrowly targeted intervention, greater focus on the evidence might have led the design to be less comprehensive and more concentrated on activities with a greater chance of success.

**There were concerning limitations in the MA's budgeting, financial, and M&E reporting systems.** Over the course of the programme, the MA moved from having comprehensive annual budgets broken down by meaningful categories, to having more of a rolling workplan and cash-budgeting approach. Furthermore, the coding system used in the budgets for decentralised funds did not correspond to the coding in the expenditure data. If budgets are not prepared and agreed over a meaningful timeframe (such as a year or a tranche), and do not have the same categorisation system as expenditure tracking, there is no framework to ensure that expenditure is guided by the objectives of the programme, or to assess whether expenditure and implementation is on track. Transparency around the budgets and expenditure tracking, at least to a wider annual review audience, diminished over the course of the programme. Furthermore, the difficulty in receiving a reliable dataset for LGA expenditure is not a sign of a well-functioning financial management system. In addition, the MA's M&E system also displayed increasing weaknesses over time. The annual monitoring surveys contained appropriate quality control mechanisms, but the periodic Fact Sheets lacked quality-control processes. The logframe indicators changed substantially year on year, and the MA's Annual Reports became increasingly narrow, limiting the use of both of these as monitoring tools to hold the programme to account. (4.6, 4.7)

## Sustainability

**Finally, many of the elements necessary for sustainability of parts of the EQUIP-T programme are in place, but not all.** Some main activities, such as teacher in-service training, training SCs to engage the community in school improvement, and holding DEMs, are seen as effective by stakeholders at the school and district level. Furthermore, Tanzania has the organisational and institutional capacity to continue these activities: the knowledge is already embedded in the government system. Sustaining these activities will need continued strong leadership and messaging from the top of government, right through the levels of administration, with guidance and follow-up to make sure they become institutionalised. However, affordability is an issue, with all activities having at least an opportunity cost resulting from the time spent by public sector workers attending to these activities rather than other duties, and in some cases substantial budgetary costs of, for example,

<sup>61</sup> The MA points out that decisions to prioritise construction and delay gender and inclusion activities were taken to ensure alignment with government priorities; these decisions were not made lightly and were taken in the full knowledge that these actions would negatively impact on programme results in the short term.

holding training sessions. Many activities may appear ‘zero-cost’ but realistically have some minimal cost, and this is not currently affordable in LGAs’ existing budgets. (5.3, 7.3, 8.3)

## 9.2 Recommendations

The recommendations below are intended to guide the Government of Tanzania and DFID in their adoption and adaptation of activities from EQUIP-T, as well as to identify broader areas where action would be valuable for improving equitable access to high-quality education at the primary level in Tanzania. The recommendations for government are particularly important given the government’s commitment to sustain six elements of EQUIP-T: the in-service training model; SRP; SIS; PTPs; IGA; and M&E. Meanwhile, DFID is planning its next education support programme in Tanzania (called Shule Bora) now that EQUIP-T is coming to an end; thus, the recommendations directly for DFID are broadly about the new programme design, but also may be relevant for future programmes. Further, the recommendations for government are relevant for DFID too, in considering the details of the objectives and activities of the new programme. In addition, there are other programmes being implemented or developed (such as the government’s lead in shaping the new GPE LANES 2 programme) which would benefit from these same recommendations.

### 9.2.1 Recommendations for the Government of Tanzania

The recommendations for government are also repeated in Annex G, which also gives an indication of the responsible institutions, and level of priority recommended by the evaluation team.

#### Component 1: Improving teacher performance

1. **Under the NF-TCPD, continue the in-service training model in EQUIP-T regions and scale this up to other regions that do not already have an effective school-based model.** The EQUIP-T in-service training model has been effective in improving teacher confidence and morale, and in empowering teachers with a range of approaches which have contributed to improved learning outcomes. The model conforms to the principles and modalities set out in the NF-TCPD, and the numeracy, literacy, and other modules have been approved by TIE.<sup>62</sup>

The model should continue at school level, which is lower cost, effective at providing ongoing support, and means teachers receive training despite high turnover. However, there does need to be intermittent face-time with experts (such as TTC tutors or TRC staff) to refresh key ideas and troubleshoot when there are problems. The revamping of TRCs, under the NF-TCPD, will help to support this need.

The school-based model should be institutionalised through the formal recognition of the in-service coordinator role in schools (which may need promotion and additional salary), and by building this into the monitoring and quality assurance responsibilities of WEOs and SQAs, as laid out in the NF-TCPD.

The participation and completion of in-service training should be built into the teacher career progression framework to incentivise taking part. The NF-TCPD states the intention to recognise and certify TCPD based on participation in standard modules that are accessible to all teachers.<sup>63</sup>

<sup>62</sup> The NF-CPD (p. 9) states that TCPD should be grounded in collaborative, inclusive, gender-responsive, and participatory learning. It also emphasises the importance of non-residential models which allow teachers to be on task during school hours (p. 22). These elements are all central to EQUIP-T’s model.

<sup>63</sup> See MoEST (2019).

The best practices from the in-service training content should be integrated into pre-service training programmes.

At the time of this evaluation's fieldwork, it was clear that teachers faced challenges in managing large classes. The 'general effective pedagogy' module, rolled out by EQUIP-T after the qualitative fieldwork, will have been important in providing content on methods for large classes.

The government will need to increase budget allocation to education, including through LGAs, to sustain the in-service training model.

2. **Make use of the TLMs distributed under EQUIP-T.** The reading books are not currently being used, and they are a valuable resource. The government could include lessons on how to use reading books into both pre-service and in-service training, and include use of the books in curriculum guidelines (such as a minimum number of minutes per week for pupil reading practice).
3. **Review, with partners, the support for teachers on teaching children who do not speak Kiswahili as their mother tongue.** Many teachers are in contexts where many pupils do not speak Kiswahili as their mother tongue. This evaluation shows that these pupils are far behind their classmates in Kiswahili and maths, and wider evidence shows it could take many years for these pupils to catch up (Collier, 1989). Teachers would benefit from support in acquiring skills in approaches to teaching pupils in multilingual classrooms, and in putting these into practice. Such techniques could include games, group work, and use of translanguaging<sup>64</sup> (Heugh et al., 2019). Interventions to help children catch-up before starting, such as the SRP, may also be beneficial, and this would be a useful focus for future research in Tanzania.
4. **Continue the focus on reducing classroom shortages and recruiting more teachers.** The evaluation confirms the extremely large pupil-to-classroom and pupil-to-teacher ratios and the challenges faced by teachers in this context. Building more infrastructure and recruiting more teachers are necessary to alleviate these barriers to effective teaching.
5. **Review the teacher management policies which lead to high absenteeism and turnover.** The extremely high level of teacher absenteeism and turnover affects the usefulness of any intervention to improve teaching quality in the classroom. Improved monitoring and accountability for classroom attendance, and reducing the rate of transfers, would help address this. The recommendation to review transfer and turnover is also relevant to head teachers, WEOs, and LGA officers.

## Component 2: Strengthened SLM

6. **If the government is planning to continue using SIS, it needs improvements from the version seen by the evaluation team.** It should be reviewed in terms of ease of data entry for school staff and how to make the data useful at the school level. The SIS should not replace existing data collection systems until it is proven to work and all glitches have been removed for a whole annual cycle. More than one person per school should be trained in how to enter and use the data.

## Component 3: Strengthened district planning and management

7. **Continue and scale-up DEMs across the country.** The introduction of DEMs has been successful at improving relationships between WEOs and the education department, sharing experience and learning, and allowing more efficient information sharing and management of schools. Ideally, DEMs should include an aspect of demand-driven training (akin to WEO CPD). Based on the findings of the evaluation, this roll-out will need to involve some training, rather than

---

<sup>64</sup> Translanguaging includes a range of processes in which bi-/multilingual people make use of the knowledge they have of many languages and how to use these languages. This can include alternating between two or more languages.

just self-reading materials, with examples of best practice, and continued leadership and emphasis from all levels of government to maintain and institutionalise the practices.

8. **Produce a standardised manual on WEOs' roles and roll these out through DEMs.** WEOs lack previous training on their roles and responsibilities, and appreciate the training they have received under EQUIP-T; this suggests the gap is likely to apply in the rest of the country and that a manual would be useful for new WEOs at least. After initial roll-out, the manual could be used within ongoing DEMs to discuss how WEO responsibilities should be performed and to guide the CPD aspect of these meetings.
9. **Continue providing responsibility allowances for WEOs and strengthen accountability for their performance.** The experience of the WEO grant under EQUIP-T has been that it unlocked more frequent visits to schools by WEOs. The government introduced a responsibility allowance in 2016, and although WEOs in EQUIP-T regions will now lose the value of the EQUIP-T grant, they will still have more resources than in the past, which is likely to make a substantial contribution to WEOs' performance. Although the use of the responsibility allowance is not stipulated (though perhaps should be), the government should enforce performance management of WEOs so that they are accountable for making visits to schools and supporting school improvement.

#### **Component 4: Stronger community participation and demand for accountability in education**

10. **Review how to strengthen parental engagement and school accountability to the community, drawing on the best practices of SCs and PTPs.** SCs play an important and recognised role, including in engagement with parents. The SC parent member is seen as a preferred channel for parents to raise concerns, rather than via the PTP. In many cases, introducing PTPs has caused confusion and minimal additional benefit in terms of wider parental engagement and school accountability. However, some PTPs are carrying out activities which are seen as beneficial within the school, even if they are not recognised as bringing the wider parental body closer to the school. It would be worthwhile considering options for encouraging the continuation and scale-up of these types of parent-led activities – involving PTPs and/or SCs. Future efforts to strengthen wider parental engagement and school accountability should recognise the central role of the SC, as well taking into account the constraints to participation and empowerment faced by parents, in deciding how to engage and represent them.
11. **Consider continuing distribution of school grants for IGA or school improvement.** The PTP grants and IGA grant have energised community mobilisation and participation, in terms of labour, in-kind resources, and interest in the school. The PTP grants allowed direct school improvement and gender welfare activities, whereas IGA projects, if successful, create an ongoing income stream for the school. The PTP grant does not require a PTP to be successful – the same aim could be achieved through the SC – and thus this recommendation can be considered separately. The cost analysis shows that the cost of these grants is actually rather small in comparison with current LGA budgets. Distributing grants for IGA or school improvement – in new regions which have not received them already – would help the SC to engage more with the school and the wider community.

#### **The sustainability of best practices**

12. **Embed any new practices throughout the management chain.** Where activities are continued or scaled up to other regions, they need to be embedded in the responsibilities and monitoring systems throughout the chain – in how head teachers supervise teachers, how WEOs and SQAs supervise and help ensure quality in schools, how LGAs monitor WEOs, right up to regions and central ministries. At LGA level, DEMs are a platform to collect information on practices from the school and WEO level (such as in-service training). Regions should continue collecting performance and activity reports from LGAs as this accountability has contributed to an increased focus on results.

13. Recent trends in the education sector budget suggest that paying for the three activities studied here – teacher in-service training, PTP grants, and IGA grants with initial training – will not be affordable for the LGAs without additional financial support. The central government should take two actions if it wishes to continue replicating and scaling-up these activities across the country.
- a. **Recognise the cost burden and provide sufficient budget for these activities**, whether that budget is held and spent at national, regional, LGA, or school level. Government should not expect LGAs and schools to implement these activities if no provision is made for the costs. Government should also consider ways to reduce costs whilst maintaining satisfactory quality, as reviewed in the cost study (OPM 2020b).
  - b. **MoEST and PO-RALG should strengthen the case for additional spending in the education sector**, to put to MoFP and Parliament. This requires reviewing and assessing the evidence, and communicating this evidence to MoFP, the Cabinet, and parliamentarians, who each have a role in approving the final budget. Furthermore, MoFP should present annual budget allocation and disbursements at the joint annual sector review to provide additional accountability.

### 9.2.2 Recommendations for DFID

1. **An evidence-informed TOC should be more explicit in the initial design of the programme, and it should be used for learning and adapting during implementation.** Setting out the assumptions more clearly at the start of the programme will allow more realistic reflection on the risks for programme success, and what needs to be monitored and tested before scaling. If DFID's new programme intends to 'test, adapt, and deliver at scale', then the learning process for adapting needs to be systematic, with critical reflection, adjustments, and documentation of those changes and the reasoning behind them. DFID should require this type of reflection and documentation as part of its annual reviews. The focus on evidence-based TOCs in the draft TOR for Shule Bora is promising (DFID, 2019).
2. **Allow more room for discretion in the allocation of decentralised funds by the LGAs.** Whilst the mitigation of fiduciary risk should not be compromised, if LGAs had some room for decision making in how to use the funds (even within certain options/boundaries), this would increase local ownership, improve targeting towards local needs, and allow LGAs to practice planning and budgeting skills.
3. **Programmes should have budgets for medium-term periods (such as annual) which are transparently reported against, and robust financial tracking systems.** For future programmes, DFID should have closer sight in agreeing and monitoring progress against budgets for meaningful periods (such as annual) and component-level budgets. There should always be room for iterations as the context and programme change, but there should be agreed budgets to guide these iterations and explain changes. Budgets and expenditure need to be categorised in the same way to allow comparison. Expenditure tracking should be quality controlled and reliable. Expenditure and budget execution should be reported transparently in annual reviews.
4. **Ensure the financial system is set up to aid monitoring and accountability, and the assessment of value for money.** For example:
  - a. The activity coding structure should strictly relate to sub-component categories.
  - b. Put in place a level of classification in financial data which relates to whether the activity was an overhead/development cost, or an implementation cost.
  - c. The category for implementation costs should further be coded by which region it is for, or if it is across all regions.
5. **Monitoring data should be comprehensive, regularly updated, quality assured, and ideally track actual beneficiaries.** Data should be collected and stored using a database. There should be documented standards and processes for ensuring data quality, and a sample of this data

should be verified on at least an annual basis. Also at the design stage, consider how the monitoring data can be made compatible with the coding of financial data. This monitoring system needs sufficient resourcing. The emphasis on bottom-up and comprehensive monitoring in the draft TOR for Shule Bora is positive.

6. **DFID should consider how to support the government in implementing the recommendations made above.** This would include integrating these recommendations into the design of DFID's new programme to support the government with the activities with highest impact. The programme could both sustain effective activities in the nine EQUIP-T regions and incentivise activities for national scale-up through the results-based financing component.

### 9.3 Lessons for future programmes in Tanzania and other countries

Various lessons of potential wider relevance for the design and implementation of education policies and programmes can be identified from the evaluation:

1. **Where teaching quality is low, and teachers lack access to regular training opportunities, the provision of teacher in-service training focusing on pedagogical practices is likely to improve learning outcomes.** However, the context is also important, and there will be more and less opportune times for such a training to be introduced if it is to be well-received. In the case of EQUIP-T, the change in Standards 1 and 2 curricula demanded new skills from teachers which made the training modules immediately relevant and useful for teachers.
2. **Where programmes have an objective to improve early grade teaching and learning, the existing language capabilities of pupils will be key to designing effective interventions.** Conducting baseline research to understand the language comprehension and pre-school experience would clarify whether interventions to increase school readiness and support transition to school, as well as support for teachers in these contexts, are necessary.
3. **As has been found in other studies (Glewwe and Muralidharan, 2015), the distribution of TLMs does not guarantee that they will be used.** In this case, most schools had received the supplementary reading books for early grade pupils but were not using them. Efforts to increase the use of TLMs will need to go further, to understand what makes teachers use them or not. This evaluation did not research why the materials were not being used, but factors which might be important for TLM distribution in general include: better sensitisation for teachers on what the TLMs are for and how to use them (integrated into the in-service training content); curriculum directives to use the TLMs; ensuring there are sufficient quantities for every child to be able to access the TLMs (so a slight shortage does not stop all children accessing the materials); reassuring teachers that they should be used and it is OK that the materials will get worn; and sensitising supervisors (head teachers, WEOs, and SQAs) on how to use the TLMs and on their roles in monitoring and supporting their use.
4. **The design of capacity building programmes in contexts with major system-wide constraints needs to take these into account and be cognisant of the risks to successful delivery and outcomes.** In particular, issues of high staff turnover, staff absenteeism, and staff shortages affect how the new capacity is put into use and retained in the system. As with EQUIP-T, refresher training is one way to mitigate these challenges, and programmes should consider whether more is needed.
5. **Increasing parental engagement and representation in school matters is very challenging, and is not ever likely to be quick or easy to change.** The cultural barriers and relationships are deeply engrained, and since parents are busy, there is a high opportunity cost to attending meetings, which cannot easily (and affordably) be overcome. Interventions targeting parental engagement need careful design and piloting, as well as setting realistic ambitions for the level of change possible in the programme period.

6. **This evaluation finds that continuing many of the activities will depend on strong leadership and availability of funds, which are by no means certain.** The challenge of sustaining donor-led initiatives is universal in cases where funding comes to an end and governments are unlikely to find new domestic sources of funds to fill the gap. Programmes should assess sustainability early and support efforts to lobby for more domestic funds, whilst at the same time being realistic about the likelihood of sustainability and the degree of sustainability acceptable to stakeholders.
7. **Programmes which aim to have a widespread effect on learning outcomes via government systems do not need to tackle all system constraints at once.** Allocating more time for evidence-based design, development, and scrutiny of the TOC and regular monitoring of the TOC, as well as piloting, would be helpful, rather than using resources to implement immediately at scale. The information from monitoring and piloting should be used for learning, which may lead to stopping activities showing poor signs of effectiveness.<sup>65</sup>
8. **Programmes which are designed to be adaptive would benefit from an evaluation system which is tailored and responsive, in order to provide accountability.** An evaluation design which intends to quantify impact is ideally best suited to programmes which faithfully implement their original design. Furthermore, in adaptive programmes, the formal reporting for accountability purposes (such as through logframes) should have more emphasis on the learning process of the programme, rather than highly specific outputs, which are likely to change regularly.
9. **The measurement of value-for-money criteria is dependent on having a financial system set up to aid monitoring and accountability, as well as having reliable monitoring data which can be linked to financial inputs.** Budgets and expenditure should be categorised and coded so that the activity or sub-component is clear (with no overlaps), the geographical unit is clear, and the cost can be identified as a development/set-up cost or a recurring implementation cost.

---

<sup>65</sup> The draft TOR for Shule Bora has a promising focus on choosing the appropriate scale for implementation based on evidence, as well as monitoring and learning to adapt interventions.

## References

- 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.
- Collier, V. P. (1989) 'How Long? A Synthesis of Research on Academic Achievement in a Second Language', *TESOL Quarterly*. Vol. 23, No. 3, pp. 509–531.
- DFID (2011) *DFID Ethics Principles for Research and Evaluation*.
- DFID (2019) Shule Bora (Quality School) UKAid. Terms of Reference (ToR) for Managing Agent Contract [draft as at December 2019].
- EQUIP-T MA (2014) Inception Report. Draft Final, 31 January 2014.
- EQUIP-T (2015a) Ruzuku ya Ushirikiano wa Wazazi na Walimu. Kiongozi Cha Mwezesaji (PTP grant Trainers Guide, Final version).
- EQUIP-T MA (2015b) EQUIP-Tanzania Annual Report 2015. United Republic of Tanzania, PMO-RALG, MOEVT, and DFID.
- EQUIP-T MA (2015c) Summary of Component 4; Annex 89 to EQUIP-Tanzania Annual Report.
- EQUIP-T MA (2016) EQUIP Tanzania, 2016 Annual Report External Version. United Republic of Tanzania, DFID.
- EQUIP-T MA (2017a) EQUIP-Tanzania Proposed Programme Extension. (Extension proposal as at 17 February 2017).
- EQUIP-T MA (2017b) EQUIP Tanzania, 2016–17 Annual Summary External Version. URT, DFID.
- EQUIP-T MA (2018) Report on Parent Teacher Partnership Grant Implementation. Key Document 15 in 2018 Annual Report.
- EQUIP-T MA (2019) Quarterly Report 22. Period: April to June 2019.
- EQUIP-T MA (no date, 2019 or 2020) Strengthening Parental Engagement in Schools: Parent Teacher Partnerships. EQUIP-Tanzania Practice Paper.
- 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.
- Greene, J. C., Caracelli, V. J., and Graham, W. F., (1989) 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational Evaluation and Policy Analysis*. Vol. 11, No. 3. pp.255–274.
- Heugh, K., French, M., Armitage, J., Taylor-Leech, K., Billingham, N., and Ollerhead, S. (2019) *Using multilingual approaches: moving from theory to practice. A resource book of strategies, activities and projects for the classroom*. British Council.
- IPSOS (2015) EQUIP-T Annual Monitoring Survey. Draft Final Report. Annex 61 of EQUIP-T MA 2015 Annual Report.
- KPMG (2019) Fiduciary Risk Assessment of the Education Sector in Tanzania – 2019. Final report, 31 May 2019.
- Levin, H.M. and McEwan, P.J. (2001) *Cost Effectiveness Analysis: Methods and Applications*. 2nd Edition, Sage Publications Inc.

- Lieberman, E. S. and Zhou, Y. (2019) Validated Participation Promotes Self-Efficacy and Citizen Engagement in Development (14 September 14). MIT Political Science Department Research Paper No. 2018-2.
- MoEST and PO-RALG (2018) Tanzania Education Program for Results (EPforR) Annual Report 2017–18, Draft Approved for Circulation (9 October 2018). United Republic of Tanzania, Government of Tanzania.
- MoEST (2019) National Framework for Teachers Continuous Professional Development. United Republic of Tanzania, Government of Tanzania.
- 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]
- NECTA (2018) The Analysis of the Assessment of Reading, Writing and Arithmetic (3Rs) of Standard 2 pupils 2017. United Republic of Tanzania, Government of Tanzania.
- 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]
- OPM (2015a) EQUIP-Tanzania IE Final Baseline Technical Report, Volume I: Results and Discussion. OPM.
- OPM (2015a) EQUIP-Tanzania IE Final Baseline Technical Report, Volume I: Results and Discussion. OPM.
- OPM (2016) EQUIP-Tanzania IE: Midline Planning Report. OPM.
- OPM (2017a) EQUIP-Tanzania IE Final Midline Technical Report, Volume I: Results and Discussion. OPM.
- OPM (2017b) EQUIP-Tanzania IE Final Midline Technical Report, Volume II: Methods and Supplementary Evidence. OPM.
- OPM (2018) EQUIP-Tanzania IE. Endline Planning Report: Part I. Quantitative research.
- OPM (2019a) EQUIP-Tanzania IE. Endline Quantitative Technical Report, Volume I. Results and Discussion.
- OPM (2019b) EQUIP-Tanzania IE. Endline Planning Report: Part II. Qualitative research, cost study, and endline synthesis.
- OPM (2019c) EQUIP-Tanzania IE. Endline Quantitative Technical Report, Volume II. Methods and Supplementary Evidence.
- OPM (2020a) EQUIP-Tanzania IE. Endline Qualitative Technical Report.
- OPM (2020b) EQUIP-Tanzania IE. Cost Study.
- OPM (2020c) Background Note to the EQUIP-T IE: Observations on the management of EQUIP-T. (Internal note for DFID).
- PO-RALG and MoEVT (2015) School Committee Training Guide. Education Quality Improvement Programme in Tanzania. Funded by DFID.
- PO-RALG (2019) RE: Sustainability of EQUIP-Tanzania Best Practices. Letter from PO-RALG to EQUIP-T National Coordinator, cc. DFID. 16 January 2019.
- 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]

Uwezo (2013) Are Our Children Learning? Annual Learning Assessment Report 2012. Uwezo Tanzania.

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

World Bank (2015) Tanzania 2014 Service Delivery Indicators: Technical Report – Education Sector. Draft of April 2015.

## 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.<sup>66</sup>

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:

1. "Assess if the EQUIP-T interventions in supported councils [districts] contribute to better basic learning outcomes amongst primary school age students."; and
2. 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.<sup>67</sup>

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 of

---

<sup>66</sup> DFID-OPM correspondence March 10, 2014.

<sup>67</sup> OPM's technical proposal is available on request.

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 theory of change** 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 theory of change 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 theory of change including contextual factors.

The EQUIP-T programme theory of change 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 theory of 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 8, together with what the impact evaluation will measure, given the EQUIP-T programme design and changes to the TOR agreed with DFID.

**Table 8. 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 <sup>68</sup> .
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 MOEVT<sup>69</sup> 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 evaluation<sup>70</sup>.

- Scope of IE 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);
- To obtain school sample: in the second stage, match control schools to treatment schools using PSM instead of random selection;

<sup>68</sup> MA-OPM and DFID-OPM correspondence December 2, 2013.

<sup>69</sup> 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).

<sup>70</sup> The draft inception report was subsequently further revised based on comments received from the SEQAS review on March 7, 2014.

- 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 Final Endline evaluation questions

The following tables set out the detailed evaluation questions, sub-questions and research questions developed and answered as the basis for this final endline report.

**Table 9. Programme level evaluation questions**

Evaluation Questions	Evaluation Sub-questions	Research questions	Sources
To what extent was the programme relevant?			
To what extent was the programme relevant?	To what extent was the programme appropriately designed to achieve its results?		• Baseline report
What impact has the programme had?			
What impact has the programme had?	What has been the impact of the programme on pupil learning outcomes?		• Quantitative report
To what extent was the programme effective?			
To what extent has the programme achieved its objectives?	Were targets for improved learning outcomes achieved? Did learning gaps narrow for marginalised groups (girls, non-Kiswahili speakers)?		• Quantitative report
What factors explain these results?	What explains the changes in learning outcomes?	What explains the impact results? What explains the trends in results?	• Quantitative report • Qualitative report
	Did the demand-side assumptions required for the programme to be effective hold?	Did assumptions on pupil attendance, pre-school attendance, support at home, pupils' household poverty status, hold? Have there been changes over time?	• Quantitative report • Qualitative report
To what extent was the programme efficient?			
To what extent was the programme cost-effective?	What has the programme cost to deliver, and what are the main cost elements?		• Cost Study
	How cost efficient has the programme been in delivering (selected) outputs? What can be said about cost-effectiveness of the programme?		• Cost Study
To what extent was the programme implemented in line with its agreed budget and timetable?	To what extent were adequate and complete budgets prepared and used to guide expenditure?		• Cost Study
	To what extent were budgets fully executed, and how did the rate of budget execution vary between categories/activities?		• Cost Study
How well did key	Was evidence, and a theory of		• Background

management arrangements support programme implementation?	change, use to guide the programme's design and adaptation?		note on Management
	Were strong M&E systems in place to monitor and adapt implementation, and provide accountability?		• Background note on Management
	Were strong financial systems in place to monitor and spending, and provide accountability?		• Background note on Management

**Table 10. Component 1 evaluation questions: Improved access to quality education**

Evaluation Questions	Evaluation Sub-questions	Research questions	Sources
To what extent was the component effective?			
To what extent did the component achieve its intermediate outcome objectives?	Did the programme achieve its outcome targets for teacher performance?	Has classroom teaching improved (more active, inclusive, gender-responsive & use of TLM)? Are teachers using more positive behaviour management? Have instructional hours increased? What aspects or elements of the EQUIP-T in-service training do teachers find most useful and why? How have teachers translated learnings from in-service training to practice in the classroom?	<ul style="list-style-type: none"> <li>• Quantitative report</li> <li>• Qualitative report</li> </ul>
What factors explain the results?	Were inputs (residential and ward/cluster training, TLMs) delivered as intended?	Did EG teachers receive district/ward INSET as intended? Did school-based INSET happen as intended? Did school receive materials as intended?	<ul style="list-style-type: none"> <li>• Quantitative report</li> <li>• Qualitative report</li> </ul>
	Were outputs of the teacher component achieved as intended? If not, why not?	Did EG teachers' capacity (curriculum knowledge, general/inclusive pedagogical skills) increase? Are teachers more confident? Are materials in classrooms? Are teacher COL structures operating? Did assumptions on INSET attendance hold? What challenges, if any, do teachers face in attending and learning from INSET?	<ul style="list-style-type: none"> <li>• Quantitative report</li> <li>• Qualitative report</li> </ul>
	Did the assumptions required for outputs to lead to outcomes for the teacher component hold?	Did assumptions on school 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?	<ul style="list-style-type: none"> <li>• Quantitative report</li> <li>• Qualitative report</li> </ul>

		What are the biggest challenges that teachers continue to face in their teaching practice? Do they perceive in-service training to be adequate in addressing these challenges – why so?	
	What unanticipated and unavoidable factors affected the results?		• Quantitative report
To what extent is the component sustainable?			
To what extent will the results of the component continue after the programme ends?	Will teachers continue to use the practices they have learnt, when EQUIP-T ends?	Will the teachers who receive the training continue to use the skills they have received? And the materials? Will WEOs/Heads continue to support teachers using what they've learnt?	• Qualitative report
To what extent are factors favouring sustainability of the teacher professional development model in place?	Was the teacher component considered effective? Are key stakeholders committed to continuing these activities and practices? Is the capacity in place to continue the activities?	Will schools continue holding school based INSET? Is the commitment, capacity and resource there at LGA level to continue?	• Quantitative report • Qualitative report
	Is it affordable for the government to continue to replicate and scale up the in-service training model?		• Cost study

**Table 11. Component 2 evaluation questions: Strengthened school leadership and management**

Evaluation Questions	Evaluation Sub-questions	Research questions	Sources
To what extent was the component effective?			
To what extent did the component achieve its intermediate outcome objectives?	Did the SLM component achieve its targets for school leadership and management?	Has SLM improved? In which areas? What actions has the head teacher taken to improve the school? Have some activities from SDPs been implemented? Does the head teacher use the SIS for SLM? Does the head teacher make public key information on pupil performance, teacher/pupil attendance & resource use?	• Quantitative report
What factors explain the results?	Were inputs (SLM training, tablets) delivered as intended?	Did head teachers receive training as intended? Did schools receive tablets?	• Quantitative report
	Were outputs of the SLM component achieved as intended? If not, why not?	Has head teacher SLM capacity increased in core areas (planning, tch & resource management, community engagement, communication/ reporting)? Are 'quality' SDPs available for the current year (evidence-based, costed, incl. teaching and learning & inclusive education activities)? Were INSET materials	• Quantitative report

		relevant/accessible? Are SPMMs happening? How frequently? What is discussed? Are head teachers meeting as COLs? If SPMMs or COLs are not being attended, why not? Is there low turnover of head teachers and WEOs? Is the SIS up-to-date with pupil/teacher records? Is daily attendance being recorded? Can head teachers produce useful analysis/reports from the SIS?	
	Did the assumptions required for outputs to lead to outcomes for the SLM component hold?	Is SIS software and hardware well designed for purpose? Did assumptions on attendance, punctuality, turnover, morale/motivation hold? If not, why? Have there been changes over time? Are supporting conditions for head teachers (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?	• Quantitative report
	What unanticipated and unavoidable factors affected the results?		• Quantitative report
To what extent is the component sustainable?			
To what extent are factors favouring sustainability in place?	To what extent are factors favouring sustainability in place?	Will head teachers/schools carry on using the skills and practices they have learnt? Will head teachers continue using SIS?	• Quantitative report

**Table 12. Component 3 evaluation questions: Strengthened district planning and management**

Evaluation Questions	Evaluation Sub-questions	Research questions	Sources
To what extent was the component effective?			
To what extent did the component achieve its intermediate outcome objectives?	Did the district component achieve its outcome targets for district planning and management?	To what extent are districts making better use of resources? Is there increased funding to education within districts? Were EQUIP-T LGA grant activities implemented as planned? Do LGAs make regular school visits and use the information? Are DEMs used to manage education issues?	• Qualitative report
		Are WEOs providing better monitoring and advisory support to all schools?	• Quantitative report • Qualitative report
What factors	Were inputs delivered as	Did LGA officers receive training? Did	• Qualitative

explain the results?	intended?	they receive grants, including grants for monitoring schools? Did they receive mentoring support? Did WEOs receive motorbikes and grants? Did they receive training under other components and on roles and responsibilities? Was initial training on WEO CPD held?	report
	Were outputs of the district component achieved as intended? If not, why not?	To what extent have district education offices' capacity in PFM improved? Are LGAs holding and attending DEMs? Was the training relevant and accessible? Were grants timely and released in full?	• Qualitative report
		To what extent has WEOs' capacity improved? Are WEOs visiting schools more regularly? Are WEOs attending DEMs and WEO CPD? To what extent are these useful? Do DEMs and WEO CPD take place without payments?	• Quantitative report • Qualitative report
	Did the assumptions required for outputs to lead to outcomes for the district component hold?	Were LGAs' overall budgets maintained? Were there external interruptions to the budget? Were funds mismanaged? Is WEOs motivation high? Is staff turnover low?	• Qualitative report
	What unanticipated and unavoidable factors affected the results?		• Quantitative report • Qualitative report
To what extent is the component sustainable?			
To what extent are factors favouring sustainability in place?	Was the component considered effective? Are key stakeholders committed to continuing these activities and practices? Is the capacity in place to continue the activities?	What practices would LGA officers/WEOs like to continue? What practices do they think will continue, and why? How likely is it that districts will continue to hold DEMs and WEO CPD, and WEOs attend? Will WEOs continue to visit schools regularly? Are key stakeholders committed to continuing improved practices?	• Qualitative report

**Table 13. Component 4 evaluation questions: Stronger community participation and demand for accountability in education**

Evaluation Questions	Evaluation Sub-questions	Research questions	Sources
To what extent was the component effective?			
To what extent did the component achieve its intermediate outcome objectives?	Did the community component achieve its outcome targets for community engagement?	Is there better engagement between schools and all parents/ the community? Do parents receive regular information on the progress of their children? Do parents receive information in other	• Quantitative report • Qualitative report

		ways? What is the perceived impact of PTPs, PTP grants, the SC training and IGA?	
What factors explain the results?	Were inputs (training on SCs, PTPs, and grants; PTP and IGA grants, noticeboards, CENAs, Shujaaz magazine and JUU club) delivered as intended?	Was training received as intended? Were training materials relevant and accessible? Did the cascade model of training work for SCs and PTPs? Did schools receive PTP and IGA grants? Did schools receive noticeboards? Were CENAs conducted?	<ul style="list-style-type: none"> <li>• Quantitative report</li> <li>• Qualitative report</li> </ul>
		Did JUU club training take place? Did schools receive the grant? Did school receive Shujaaz magazines?	<ul style="list-style-type: none"> <li>• Quantitative report</li> </ul>
	Were outputs of the community component achieved as intended? If not, why not?	Did SC capacity increase? Do SCs meet regularly? What do SCs discuss? Were PTPs set up as intended? Are SCs and PTPs more active, and are parents and the community engaged in school improvement? How did the school and community work together to spend the PTP grants and the IGAs? What has been their experience of planning for and spending these grants? Do schools have publicly accessible noticeboards, and what is displayed on them? Are head teachers aware of CENA action plans, or of any activities resulting from the CENA action plan?	<ul style="list-style-type: none"> <li>• Quantitative report</li> <li>• Qualitative report</li> </ul>
		Do schools have JUU clubs? Do they meet regularly? What actions have the JUU club taken in the last school year? Has the second PTP grant been spent and if so, what on?	<ul style="list-style-type: none"> <li>• Quantitative report</li> </ul>
	Did the assumptions hold?	Do SC/PTP members and parents have time, resources & motivation to change/engage? Are all parents aware of PTPs? Are PTPs able to self-organise? How did the school and community work together to increase awareness of the PTP, increase parental involvement and decision-making regarding actions to improve education outcomes in the schools? Is the role of the SC and PTP clear and distinct? Is there no conflict between SC & PTPs? Do parents read the school noticeboard?	<ul style="list-style-type: none"> <li>• Quantitative report</li> <li>• Qualitative report</li> </ul>
What unanticipated and unavoidable factors affected		<ul style="list-style-type: none"> <li>• Quantitative</li> </ul>	

	the results?		report • Qualitative report
To what extent is the component sustainable?			
To what extent are factors favouring sustainability in place?	Was the component considered effective? Are key stakeholders committed to continuing these activities and practices? Is the capacity in place to continue the activities?	Will PTPs and SCs remain engaged? Will schools carry on IGAs? Do stakeholders want these to continue, and what do they think needs to be in place for it to actually continue?	• Quantitative report • Qualitative report
	Is it affordable for the government to continue with PTPs and IGA model and/or scale it up?		• Cost study

## Annex C EQUIP-T Logframe, September 2019

The following are screen shots of the EQUIP-T Logframe, September 2019.

PROJECT NAME	Education Quality Improvement Programme in Tanzania (EQUIP-Tanzania)											
Better learning outcomes, especially for girls, across Tanzania	<b>Impact Indicator 1</b>		<b>Baseline 2014</b>	<b>Mid-Term 2016</b>		<b>Milestone 2017</b>		<b>Milestone 2018</b>		<b>Target 2019</b>		
	1A: % literacy and numeracy at Standard II level by Standard III. National and disaggregated by sex and also by EQUIP-T supported regions	<b>Planned</b>			EQUIP-Tanzania target regions: Literacy: 18% Numeracy: 8%	F:18% M:18% F: 8% M: 8%				EQUIP-Tanzania 5 original target regions: Literacy: 25% F:25% M:25%	Numeracy: 10% F:10% M:10%	
		<b>Achieved</b>		EQUIP-Tanzania target regions: Literacy: 12% 43,887 F: 13% 23,910 M:11% 19,998 Numeracy: 4% 14,629 F: 3% 5,518 M: 6% 10,908	EQUIP-Tanzania target regions: Literacy: 22% F: 27% M:18% Numeracy: 7% F: 6% M:9%			EQUIP-Tanzania target regions: Literacy: 18% F: 21% M:14% Numeracy: 9% F: 8% M: 11%				
	<b>Source</b>											
	1a: OPM 2014, 2016 & 2018/ 2019											
	<b>Impact Indicator 2</b>		<b>Baseline 2014</b>	<b>Mid Term 2016</b>		<b>Milestone 2017</b>		<b>Milestone 2018</b>		<b>Target 2019</b>		
	1B: National level of literacy and numeracy of Standard III children, disaggregated by sex (UWEZO 2013)	<b>Planned</b>			23%		24%		25%		26%	
		<b>Achieved</b>		Combined test: 20%								
	<b>Source</b>											
	1b: UWEZO 2013, 2016 & 2019											
<b>Impact Indicator 3</b>		<b>Baseline 2014</b>		<b>Mid-Term 2016</b>		<b>Milestone 2017</b>		<b>Milestone 2018</b>		<b>Target 2019</b>		
Standard IV Exam pass rates national, disaggregated by sex and also by EQUIP-Tanzania regions	<b>Planned</b>					<b>7 Regions</b>		<b>9 Regions</b>		<b>9 Regions</b>		
	<b>Achieved</b>			Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	
				73.38	62.77	91	90					
<b>Source</b>												
NECTA 2014, 2016 & 2019												

OUTCOME	Outcome Indicator 1	Baseline 2014		Mid-Term (2016)		Milestone 2017		Milestone 2018		Target 2019		
Improved quality of education, especially for girls, and approach ready for national scale-up	Early Primary Teacher Performance in Pedagogy: score of assessment of classroom observation of teacher methodology (sex disaggregated)	Planned			43%	Female: 43% Male: 43%	7 Regions	2 New Regions	7 Regions	2 New Regions	7 Regions	2 New Regions
		Achieved		32%	Female: 33% Male: 31%	54%	Female: 55% Male: 52%					Female 60% Male 60%
	Source											
	OPM managed Impact Evaluation 2014, 2016 & 2018/ 2019											
		Outcome Indicator 2	Baseline 2014		Mid-Term (2016)		Milestone 2017		Milestone 2018		Target 2019	
	% teachers using gender responsive pedagogy in their classroom teaching	Planned									7 Regions	2 New Regions
		Achieved				Female teachers: 70% Male teachers: 67%					Female teachers: 75% Male teachers: 70%	Female teachers: 50% Male teachers: 45%
	Source											
	Annual Monitoring Survey 2014, 2016, 2018, 2019											
		Outcome Indicator 3	Baseline 2014		Mid-Term (2016)		Milestone 2017		Milestone 2018		Target 2019	
% of standard VII girl students progressing to secondary school	Planned					7 Regions	2 New Regions	7 Regions	2 New Regions	7 Regions	2 New Regions	
	Achieved				51%	65%	70%	65%	68%	65%	74%	
Source												
Annual BEMIS 2017, 2018 & 2019												
	Outcome Indicator 4	Baseline 2014		Mid-Term 2016		Milestone 2017		Milestone 2018		Target 2019		
% and number of children (disaggregated by sex) entering Standard 1 at 8 years of age and older (National & Target Regions)	Planned		Boys	Girls	Boys	Girls	7 Regions	2 New Regions	7 Regions	2 New Regions	7 Regions	2 New Regions
	Achieved		National: 25% 168,561 EQUIP-Tanzania: 22% 45,741	National: 21% 141,274 EQUIP-Tanzania: 19% 40,531	EQUIP-Tanzania: Boys 18%, Girls 16%	EQUIP-Tanzania: 16%	EQUIP-Tanzania: Boys 15% Girls 14%	EQUIP-Tanzania: Boys 22% Girls 20%	EQUIP-Tanzania: Boys 12% Girls 12%	EQUIP-Tanzania: Boys 20% Girls 18%	EQUIP-Tanzania: Boys 10% Girls 10%	EQUIP-Tanzania: Boys 18% Girls 16%
Source												
Annual BEMIS												
	Outcome Indicator 5	Baseline 2014		Mid-Term (2016)		Milestone 2017		Milestone 2018		Target 2019		
Number of children access education through school readiness centres	Planned		Boys	Girls	Boys enrolled	Girls enrolled	Boys enrolled	Girls enrolled	Boys enrolled	Girls enrolled	Boys enrolled	Girls enrolled
	Achieved				SRP: 75 000	SRP: 75 000	A. SRP: 80 000	A. SRP: 80 000	A. SRP: 45000	A. SRP: 45000	A. SRP: 47000	A. SRP: 47000
Source												
LGA monitoring based on centres' enrollment forms												

OUTPUT 1	Output Indicator 1.1		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019			
Improved access to quality education	% schools where teachers are engaged in improving their teaching through active INSET groups					7 regions	2 new regions	7 regions	2 new regions	7 regions	2 new regions
Planned						45%	0%	60%	25%	60%	50%
Achieved						48%	0%	48.1%	38.6%	27.10%	47.10%
Source											
WEO report through LGA-based survey in 2018 &2019											
Output Indicator 1.2			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019			
Number of accessible classrooms completed						7 regions	2 new regions	7 regions	2 new regions	7 regions	2 new regions
	Planned					100 sites identified	200 accessible classrooms constructed/finished	570 accessible classrooms constructed/finished (cumulative)			
	Achieved					257 sites identified	262	741			
Source											
LGAs Engineers reports and Routine LGA monitoring											
LGA monitoring from WEO reports plus LGA quarterly visits plus EQUIP-Tanzania managing agent visits											
Output Indicator 1.3			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019			
% of schools with PTPs participating in activities to improve inclusion						7 regions	2 new regions	7 regions	2 new regions	7 regions	2 new regions
	Planned					20%	0%	70%	10%	70%	50%
	Achieved					54	0	61.0%	32.5%	37.00%	43.60%
Source											
WEO report through LGA-based survey in 2018 &2019											
Output Indicator 1.4			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019			
2017-2018: % of schools with active Juu Clubs 2019: % of schools implementing Juu clubs pack activities						7 regions	2 new regions	7 regions	2 new regions	7 regions	2 new regions
	Planned					100%	0%	95%	20%	75%	60%
	Achieved					100	0%	61.0%	45.9%	0.00%	0.00%
Source											
WEO report through LGA-based survey in 2018 &2019											
Output Indicator 1.5			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019			
Number of active running school readiness centres				7 regions	7 regions	7 regions	7 regions	2 new regions	7 regions	2 new regions	
	Planned			1000	2700	2770	2500	2500			
	Achieved			1050	2700	2770	3,051				
Source											
WEO report through LGA-based survey in 2018 &2019											
IMPACT WEIGHTING (%)	Source										
25%	WEO report through LGA-based survey in 2018 &2019										

OUTPUT 2	Output indicator 2.1		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
Enhanced school leadership and management skills	% head teachers achieving management performance indicators	Planned			7 Regions	7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
						At least 60% achieve at least 6 out 10 performance indicators		At least 60% of teachers achieve at least 7 out 10 performance indicators	At least 60% of head teachers achieve at least 3 out 10 performance indicators	At least 60% of head teachers achieve at least 8 out 10 performance indicators	At least 60% of teachers achieve at least 6 out 10 performance indicators	
		Achieved				69.0%		68.5%	97.9%	74%	97%	
Source												
WEO report through LGA-based survey in 2018 & 2019												
Output indicator 2.2			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
% of Head Teachers attending Community of Learning meetings at least twice per quarter		Planned				7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
		Achieved						54.3%	49.3%	39.60%	52.10%	
Source												
WEO report through LGA-based survey in 2018 & 2019												
Output indicator 2.3			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019				
Number of schools with at least 80% of enrolled student's baseline data already uploaded into the SIS		Planned		n/a	n/a	500 schools	1,000 schools	1,000 schools	2500 school			
		Achieved	0	n/a	107 schools	No data received	3,753 schools	3,689 schools				
Source												
SIS Database												
Output indicator 2.4			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019				
Number of Educational Managers received training on SIS data use		Planned		n/a	n/a			6,500	6,400			
		Achieved	0	n/a			6,519	5,196				
Source												
SIS Database & DED report												
Output indicator 2.5			Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
% of schools able to report through SIS to fulfill BEMIS reporting requirements					7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions
		Planned		n/a	n/a						65%	55%
		Achieved	0	n/a						86.50%	98.80%	
Source												
SIS Database & DED report												
IMPACT WEIGHTING (%)												
20%	SIS Database & DED report											

OUTPUT 3A	Output Indicator 3.1		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
Strengthened district management of education	% of WEO attending continuous professional development sessions at least twice in a quarter	Planned			7 Regions	7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
		Achieved						50%	40%	80%	80%	
								74%	94%	79%	82%	
	Source											
	Minutes/reports of CPD meetings& WEO reports											
	Output Indicator 3.2		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
	% of LGA districts with complete reconciled financial reports for EQUIP-Tanzania activities from the most recent financial quarter	Planned		30%	40%	90%	0%	95%	75%	100%	100%	
		Achieved	0	100% (changed indicator from 2015)	90% (43/48)	92%	0%	97%	100%	0%	0%	
	Source											
	Managing Agent Regional Fund Officers assessment of all LGA reports in the 7 regions & 2 New ones i.e. Katavi and Singida											
Output Indicator 3.3		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019			
Number of Districts executing District Education meetings at least four times bi annually	Planned	5 Regions	7 Regions	7 Regions	All 9 target regions		All 9 target regions		All 9 target regions			
	Achieved						50	55				
Source												
50												
53												
IMPACT WEIGHTING (%)	Source											
20%	DEO reporting based on Ward Education Officer Reports											

OUTPUT 4	Output Indicator 4.1		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
Strengthened community participation and demand for accountability in education	2018: % of schools with score card displaying updated information 2019: % of PTPs taking actions based on score card results	Planned				7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
		Achieved						50%		75%	65%	
	Source											
	WEO report through LGA-based survey in 2018 & 2019											
	WEO report through LGA-based survey in 2018 & 2019											
	Output Indicator 4.2		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
	% of schools continuing with Income Generating Activities using their IGA Grant	Planned		7 Regions	7 Regions	7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
		Achieved						80%		75%		
Source												
WEO report through LGA-based survey in 2018 & 2019												
	Output Indicator 4.3		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
	Number of schools with functional suggestion boxes in which children are raising their concerns	Planned		7 Regions	7 Regions	7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
		Achieved								120		
Source												
WEO report through LGA-based survey in 2018 & 2019												
	Output Indicator 4.4		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
	Number of community members reached with SBCC messages on girls education, transition and disability	Planned		7 Regions	7 Regions	7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
		Achieved								10,000	4,000	
Source												
WEO report through LGA-based survey in 2018 & 2019												
	Output Indicator 4.5		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
	Number of districts implemented activities to reduce Female Genital Mutilation using FGM grants	Planned		7 Regions	7 Regions	7 Regions	2 Regions	7 Regions	2 Regions	7 Regions	2 Regions	
		Achieved								16	4	
Source												
WEO report through LGA-based survey in 2018 & 2019												
	Output Indicator 4.6		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017		Milestone 2018		Overall Target 2019		
	Number of non-EQUIPT LGAs adopting key programme priority activities as a result of ISS implementation	Planned		7 Regions	7 Regions	7 Regions	2 Regions	7 Regions	2 Regions	Non EQUIP-T LGAs		
		Achieved								5		
Source												
WEO report through LGA-based survey in 2018 & 2019												

OUTPUT 5	Output Indicator 5.1		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019		
	Number of LGAs sharing learning by disseminating their own success stories	Planned				30 LGAs	50	55 LGAs		
		Achieved				45	55	51		
Source										
Learning stories disseminated and shared between LGA, with regional teams or via media and social media;DEO report										
	Output Indicator 5.2		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018		Overall Target 2019	
	Number of WEO collecting and submitting school level data through online platform	Planned					7 regions	2 regions	7 regions	2 regions
		Achieved					900	150	900	150
							835	137	965	188
Source										
DEO report & WEO monthly reports										
	Output Indicator 5.3		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019		
	Number of programme learning products for advocacy	Planned				5	7	8		
		Achieved				10	8	5		
			N/A	N/A	N/A					
Source										
Programme learning papers on website and disseminated										
	Output Indicator 5.4		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019		
	Number of published education related materials documented	Planned					Published materials	Published materials		
		Achieved					70	60		
			N/A	N/A	N/A	N/A	83	14		
Source										
DEO report										
IMPACT WEIGHTING (%)	Output Indicator 5.5		Baseline 2014	Milestone 2015	Milestone 2016	Milestone 2017	Milestone 2018	Overall Target 2019		
15%	Number of aired education related materials	Planned					Radio programmes	Radio programmes		
		Achieved					30	30		
INPUTS (£)			N/A	N/A	N/A	N/A	45	170		
INPUTS (HR)	Source									
	DEO report									

## Annex D Stakeholder engagement and impact evaluation governance

### D.1 Stakeholder engagement

Stakeholder engagement has been an ongoing process in the impact evaluation, and started from the inception phase with consultations on the overall design of the impact evaluation. Volume II of the baseline impact evaluation report sets out the stakeholder consultations carried out in the inception phase and in disseminating the baseline findings (OPM 2015b). Similarly, Volume II of the midline impact evaluation report (OPM 2017b) sets out the stakeholder engagement activities which took place between October 2015 (prior to the midline research) and December 2016 when the main dissemination of midline findings took place. Volume II of the quantitative endline report sets out the stakeholder engagement activities which took place between January 2017 and December 2018 when the main dissemination of quantitative endline took place (OPM 2019c). Plans for stakeholder engagement and dissemination of the final endline findings were set out and agreed in the Endline Planning Report part II (OPM 2019b). Table 14 summarises the main stakeholder engagement activities that have taken place since the dissemination of midline evaluation findings, and includes those planned for sharing the final endline products (this final report, the qualitative and cost study findings).

For the qualitative endline evaluation, stakeholder engagement began in January 2019 with a visit to Dar es Salaam in February 2019 to meet with DFID and the EQUIP-T MA. In February 2019, the evaluation team applied to Tanzania's Commission for Science and Technology (COSTECH) to renew approval for the endline research. The final research priorities and design for the qualitative and cost studies at endline were documented in the Endline Quantitative Planning Report part II (OPM 2019b), and this was submitted to DFID in early March 2019 and circulated to the EQUIP-T MA and the evaluation's Reference Group for feedback. This marked the end of the preparatory phase.

Draft versions of the qualitative endline report and cost study were shared with both DFID and the EQUIP-T MA for comment and correction of errors. These revised drafts were used to develop the first full draft of this endline report. After this, the evaluation team will convene a full day Reference Group meeting in Dodoma in January 2020 to present the results and to receive feedback from this wider group of stakeholders. The report will be finalised following this feedback. The evaluation team would be happy to present the findings at other forums for disseminating if invited, for example the joint sector review is particularly useful for the domestic audience.

The principal audience for this endline quantitative evaluation are DFID and GoT officials, and it will also be of interest to the remaining and former staff of EQUIP-T's MA, as well as other education stakeholders – including NGOs and development partners – in Tanzania. The results are intended promote accountability and lesson learning for DFID and the GoT. The findings will also help to inform the design of DFID's next education programme in Tanzania.

**Table 14. Stakeholder consultations and events—from dissemination of quantitative endline findings to plans for dissemination of final endline findings**

Date	Purpose
<b>Dissemination of quantitative endline findings</b>	
November 2018	Submission of draft endline quantitative evaluation report to DFID and EQUIP-T MA. Phone call meeting between evaluation team, DFID and EQUIP-T MA to discuss feedback on the draft report

December 2018	Presentation and discussion of quantitative endline findings at Reference Group meeting in Dodoma.
March 2019	Presentation of quantitative endline findings at EQUIP-T Steering Committee meeting
<b>Final endline evaluation (part II) engagement</b>	
January 2019	Consultation on the design of the qualitative endline evaluation and the cost study
February 2019	Application made to COSTECH for the 2019 endline impact evaluation studies (subsequently granted)
March 2019	Endline part II planning report (qualitative, cost and final reports) submitted to DFID, EQUIP-T MA and broader Reference Group for comment and feedback (report approved)
April 2019	Update meetings in Dar es Salaam with EQUIP-T MA and DFID on progress of the qualitative and cost studies. Interviews with EQUIP-T MA as part of qualitative study
April/May 2019	Qualitative study: research assistant training and fieldwork
July 2019	Presentation to the DFID Education Advisors cadre on findings from the quantitative endline on EQUIP-T's teacher INSET model
September 2019	Presentations at the UKFIET Oxford Conference on Education and Development on findings from the quantitative endline: 'Improving learning outcomes for marginalised children in Tanzania'  Meetings with EQUIP-T MA, PO-RALG, MoEST, and regional, district and school level officials as part of cost study. Email exchange with TIE and Tusome Pamoja for cost study.
November 2019	Submission of draft endline qualitative report and draft cost study to DFID and EQUIP-T MA  Comments and clarifications received over email from EQUIP-T MA
January 2020	Presentation and discussion of final endline report, including qualitative and cost study endline findings at Reference Group meeting in Dodoma.  Presentation of final endline key findings at EQUIP-T final dissemination event in Dodoma.  Presentation of final endline key findings and lessons to DFID Tanzania staff in Dar es Salaam.
April 2020 (planned)	Evaluation briefing note, covering key messages and recommendations. This will be intended for an audience of Tanzanian policy makers (education, finance, other sectors), programme designers and implementers to feed learning into new policies and programmes.
Post-April 2020 (planned)	The evaluation team is available to present the findings at other fora as relevant in Tanzania, or internationally. One possibility is the Joint Education Sector Review, or one of the Technical Working Groups.

All of the reports, briefing notes, issues papers and other products produced as part of the impact evaluation are available on OPM's website

A technical working paper on the innovative approach to impact estimation used in this study is also on OPM's website: <https://www.opml.co.uk/publications/working-paper-matching-differencing-repeat>

Briefing notes and conference papers produced using the impact evaluation findings have been uploaded on to the Social Science Research Network [www.ssrn.com](http://www.ssrn.com) (see for example: <https://ssrn.com/abstract=2779240>; <https://ssrn.com/abstract=2579284>; and <https://ssrn.com/abstract=2782747>).

The baseline, midline and endline quantitative survey data (anonymised) is publicly available on the World Bank microdata library <http://microdata.worldbank.org/index.php/catalog/2290>.

## D.2 Reference Group

At the start of the impact evaluation in 2014, the Ministry of Education led a process to form an EQUIP-T impact evaluation Reference Group to provide technical recommendations and feedback to the OPM evaluation team. The terms of reference for the Reference Group are included in the Midline Planning Report (OPM 2015a, Annex F). At baseline, the Reference Group held its first meeting to review and comment on the overall impact evaluation design (January 2014). A second Reference

Group meeting was held in November 2014 where baseline findings were discussed extensively, feedback provided to guide revisions to the report, and members advised the evaluation team on opportunities for dissemination as well as links with other studies and programmes. A third Reference Group meeting took place in December 2016 to discuss the draft midline evaluation report and plan for dissemination of the findings. A fourth Reference Group meeting took place in December 2018 to discuss the draft quantitative endline report and consult on the design of the qualitative and cost endline studies.

During each meeting, members provided useful feedback on the draft reports (noted in the meeting minutes, subsequently circulated to all members for corrections or additions), and members were also requested to provide any additional feedback in writing. The evaluation team consolidated all the feedback received on the draft reports from DFID, the EQUIP-T MA, and other RG members into a document. From this, the team carefully considered each comment and made changes to the draft reports as appropriate. The team also drafted a written response to each comment, explaining how the comment had been dealt with in the final report or justifying why no changes had been made. This commentary was submitted to DFID together with the final report.

From its inception, the Commissioner for Education has chaired the Reference Group and it is convened by Professor Herme Moshia from UDSM who is a core member of the evaluation team. The Reference Group membership has been refreshed in consultation with the Commissioner for Education at each key stage: in between baseline and midline, midline and quantitative endline, and now since in between the quantitative and final endline reports. This is to adjust for members no longer able to represent their organisations, and to expand to include additional education agencies and relevant perspectives to the group. The organisations represented on the Reference Group are:

- Government ministries: MoEST and PO-RALG;
- Government education departments and agencies: National Examinations Council of Tanzania (NECTA); Tanzania Institute of Education (TIE); and Agency for the Development of Education Management (ADEM).
- DFID;
- EQUIP-T MA
- UDSM, School of Education; and
- Education research organisation (Twaweza East Africa).

The Reference Group members invited to the final meeting in January 2020 are given in Table 15.

**Table 15. EQUIP-T Impact Evaluation Reference Group members (Invited January 2020)**

Member	Position	Organisation
Dr. Lyabwene Mtahabwa	Commissioner for Education	MoEST
Gerald Mweli	Deputy Permanent Secretary Education	PO-RALG
Dr. George Jidamva	Assistant Director Primary Education	PO-RALG
Neema J Chamgeni	EQUIP-T Coordinator at PO-RALG	PO-RALG
Ally Swalahe	EQUIP-T Coordinator at PO-RALG	PO-RALG
Julius Nestory	Director of Education Administration	PO-RALG
Benjamin Oganga		PO-RALG
Dr. Semakafu	Deputy Permanent Secretary	MoEST
Petro Makuru	Assistant director M&E	MoEST
Hilda Mkandawile	Senior Education officer	MoEST
Augusta Lupokela	Senior Education officer	MoEST
Pili R. Mazege	SQA officer	MoEST
Valeria Alphonse	SQA officer	MoEST

Dr. Mwambene	DPP	MoEST
Nicholaus Moshi	Principal Education Officer	MoEST
Abdalla Migila	Regional Academic Officer	REO - Dodoma
tbc	District Education Officer	DEO - Chamwino
Dr. Siston Masanja	Chief Executive Officer	ADEM - Bagomoyo
Prof. Kalafunja Osaka	Lecturer - Education department	SAUTI - University
Dr. Blackson Kanukisa	Lecturer - Education department	UDSM
Martin Chegere	Economics Department	UDSM
Johnbosco Mahundi	NECTA officer	NECTA
Dr. Aneth Komba	Director General	TIE
Fika Mwakabungu	Director for Curriculum	TIE
Dr. Godfrey Telli	Coordinator - Education	Twaweza
Ochola Wayoga	National Coordinator	TENMET
Shardul Oza	Senior Program Manager, RISE Tanzania	RISE/University of Georgetown
Kristeen Chachage	Education DPG Secretariat	Education DPG
George Senyoni	M&E specialist	EQUIP-T
Laura McInerney	Deputy National Coordinator	EQUIP-T
Vincent Katabalo	Deputy National Coordinator	EQUIP-T
John Lusingu	Education advisor	DFID
Arianna Zanolini	Education advisor	DFID

### D.3 Impact evaluation governance and quality assurance

Oversight and policy direction for the impact evaluation is provided by an OPM Governance Team comprising the OPM Managing Director, the OPM Director of Social Policy, Statistics, Evidence and Accountability, the OPM education portfolio lead, and an OPM Education Associate who is Senior Education Advisor in the impact evaluation core senior team (see Table 16 below).

Management is executed by the Project Manager, an OPM Senior Education Consultant, who in addition to playing a leading technical role is responsible for team management, the coordination of inputs, financial management and liaison with the supporting administration team and research teams in OPM's Oxford office and OPM's Tanzania Office respectively, and OPM's internal reporting and project oversight processes.

The Project Manager is responsible to the OPM Governance Team for successful delivery of the impact evaluation. The Project Manager is supported by a core senior team and a wider team of technical specialists (see Table 16 below). The core senior team comprises: the evaluation Project Director, who is an OPM Principal Education Consultant and has worked on the impact evaluation since baseline, and managed the evaluation for the midline and quantitative endline phases; the fieldwork manager who has conducted both quantitative and qualitative fieldwork, analysis and overall in-country stakeholder coordination; a senior education advisor (also part of the OPM governance team); and a senior national education advisor. There are 15 technical specialists in the wider technical team. The project manager ensures that the various teams work together to meet the objectives of the evaluation, and to produce the key deliverables. The core team is responsible for stakeholder engagement including dissemination of findings and engagement with the Reference Group.

**Table 16. Endline impact evaluation team members and roles**

Name	Role
Nicola Ruddle	Project Manager/Qualitative District Study Lead/Cost Study Lead

Georgina Rawle	Project Director/Endline Quantitative Design Lead/Quantitative Analyst
Deogardius Medardi	Quantitative Survey Fieldwork Manager / Qualitative Fieldwork Team Leader / Qualitative Analyst
Paud Murphy	Senior Education Advisor
Professor Herme Mosh	Senior National Education Advisor
Dr. Gunilla Pettersson Gelande	Senior Education Specialist/ Quantitative Analyst
Dr. Michele Binci	Impact Estimation Lead
Paul Jasper	Senior Impact Estimation Analyst
Safa Khan	Impact Estimation Analyst
Jana Harb	Quantitative Survey Fieldwork Technical Lead/ Quantitative Analyst
Ignatus Jacob	Quantitative Survey Fieldwork Lead
Andreas Kutka	Quantitative Survey Fieldwork Adviser
Diego Shirima	Quantitative Survey Data Manager
Alessio Romarri	Quantitative Research assistant
Michelle Rorich	Quantitative Research assistant
Mehjabeen Jagmag	Qualitative Lead Researcher
Kelly Casey	Qualitative Analyst
Peter Sutoris	Qualitative Fieldwork Team Leader
Gabi Elte	Cost Study Analyst
Emmanuel Maliti	Cost Study Fieldwork Advisor

Quality Assurance processes for the individual technical studies is set out in the relevant reports. For this final endline report, quality assurance has been provided using a number of layers of review. It began with consultation and advice from OPM's Principal Evaluation Consultant Stephen Jones, who also reviewed initial drafts of sub-sections of the report. The first complete draft was produced by Nicola Ruddle, and sections were internally reviewed by Georgina Rawle and Mehjabeen Jagmag.

The full draft was shared with three reviewers: Paud Murphy, Senior Education Advisor, Professor Herme Mosh (University of Dar es Salaam), Senior National Education Advisor, and Dr Caine Rolleston a leading academic researcher in the field of education and economics (Institute of Education, University College London). This team also reviewed the baseline, midline, quantitative and qualitative draft reports.

A final stage of external quality assurance will be provided through the Reference Group meeting, together with review and feedback from DFID.

Finally, this report, along with the three endline technical studies, will be reviewed by EQUALS (DFID's external review body for evaluations).

## Annex E Accordance with international best practice

### E.1 Ethical considerations

**Conducting evaluations that include disadvantaged populations require high ethical standards to ensure that expectations are not raised, confidentiality is maintained, that respondents are never forced to participate or encouraged to speak about subjects that may be traumatising, and that all activities are age appropriate.** Ethical considerations influenced the entire endline design, recruitment and management of the evaluation team; consultations and interviews with informants; data storage and use.

The evaluation design, instruments, consent forms and fieldwork protocols, for both the quantitative and qualitative rounds of research, were approved by the OPM Ethical Review Committee.

Additionally, OPM received approval for this research from the Tanzania Commission for Science and Technology (COSTECH), which has the mandate of coordinating and promoting research and technology development in the country. OPM also received approval from PO-RALG, regional and district authorities to carry out endline activities.

#### E.1.1 Principles of ethical research

As with all OPM research, the qualitative and quantitative endline studies sought to follow a set of ethical principles in conducting fieldwork based on our own experience as well as best practice standards and DFID and EU evaluation policy. A review of best practice was conducted to inform the design and protocols of the midline fieldwork and data use, and were reproduced for the endline research. This review looked at the protocols OPM used in the baseline, those used in OPM's other education evaluations, those used by other research organisation in Tanzania, and guidance from organisations specialising in children's rights (Save the Children, 1997), research (Open University, [no date], US Department of Health and Human Services, 2009) and development (DFID, 2011).

There are three basic ethical principles of research with human subjects, as set out in the Belmont Report (1979):

1. **Respect for persons:** This means the prospective participants should be given the information they need to decide whether or not they want to participate, they should be given the freedom to decide not to participate or to stop at any point.
2. **Beneficence:** This principle requires that no harm is caused by the research.
3. **Justice:** Justice requires that individuals and groups are treated fairly and equitably.

#### E.1.2 Ethical protocols in the endline research

The ethical principles above were translated into the following protocols for the research, keeping in mind that ethical principles apply to all interview respondents.

- **Informed consent:** means that potential respondents are given enough information about the research and researchers ensure that there is no explicit or implicit coercion so that potential respondents can make an informed and free decision on their possible involvement in the fieldwork. Respondents are informed that their **participation is fully voluntary**, and they can withdraw from the survey at any time. Based on our experience of research in Tanzania, we asked for verbal consent from all participants before each interview. Our experience shows that participants feel uncomfortable and sensitive about the formality of giving their signatures,

particularly government employees. It is important that participants understand what is happening, and that all information, consent statements (read orally) and instruments are translated into a language that is easily understood. Specific consent was sought from all participants before recording focus group discussions or key informant interviews.

- **Where children were interviewed** in the quantitative survey, we asked the head teacher to give consent on behalf of the parents, and also asked the children for their own consent, in simple language. The statement is read slowly, and the enumerator will read it again if necessary. This consent statement and agreement is done individually for the quantitative fieldwork, away from other teachers or parents, so that they do not feel pressured either way. If the researchers feel that any child is not comfortable during this process, they will tactfully find a way to take the child aside and discuss this with them personally.
- **Clarifying purpose:** Researchers will always clearly introduce themselves to all participants and explain, in a way that is easily understood by all, the purposes of the research and what will be done with the information provided by participants to moderate expectations regarding what participants 'gain' from joining the research. No financial compensation was provided to individual participants, but refreshments were offered during group sessions.
- **Anonymity:** given that research respondents share considerable amounts of personal information with us, it is our responsibility to ensure that their confidentiality is maintained, and personal information is protected. This is operationalised by ensuring that all datasets are anonymised, in the sense that all names of people are removed before the data is shared publicly. Furthermore, participants are interviewed in a quiet place where others cannot hear their responses. These principles are intended to avoid any social risk from views being overheard by others in the community or those above them in the reporting line, and should allow respondents to speak more honestly.
- **Ensuring the safety of participants:** this means that the environment in which research is conducted is physically safe. We also provide phone numbers for Tanzania office-based staff who the field team can contact in case of emergencies. The safety of respondents and the children fieldworkers encounter is a primary consideration and was covered under the evaluation's safeguarding policy and reporting protocol.
- **Particular care is taken in our engagement with children.** The research involves interviewing children in standard 3, who generally are between the ages of 9 and 11 years. Given their age, it is important they are treated with care and respect, and given full opportunity to decide to opt out of the work. The fieldworkers carrying out the interviews will be trained on the ethics of working with children – ensuring a safe and private space for their participation, letting them ask questions, making it clear it is fine for them to leave a question or leave the interview entirely, keeping responses confidential and anonymous – verbally but also by carefully handling the data collected. These processes are set out in the enumerator manuals which are used during training and available for reference during the fieldwork. No responses will be coerced, participants will be free to not respond.
- **Ensuring that people understand what is always happening:** This was ensured using local enumerators, to ensure that research is conducted in the appropriate language and dialect as well as ensuring that fieldworkers are familiar with local customs and terminology. Consent statements were read out in Kiswahili.
- **Minimising burden or reward:** There is no notable benefit or burden (except time) of taking part in the research, and all participants will be subject to the same benefits and burden. In the qualitative fieldwork, participants received refreshments for taking part in interviews, and respondents who were not teachers were given TZS 5,000 as transport compensation after the interview (approximately £1.50). Schools were given a box of sanitary pads for taking part in the qualitative fieldwork and a map for the quantitative fieldwork.

- All fieldworker training cover **principles of research ethics** and **respecting cultural sensitivities**. OPM's evaluations respect any differences regarding culture, local behaviours and norms, religious beliefs and practices, sexual orientation, gender roles, disability, age, ethnicity, and other social differences, such as class, when planning studies and communicating findings. We endeavour to include research participants who may be vulnerable or marginalised participants to take part in the research.

## E.2 OPM Safeguarding Policy and Reporting Protocol

In addition to ensuring this evaluation adopts the highest ethical standards, OPM put in place specific protection measures to ensure our endline qualitative research team going into schools and communities understand their ethical and statutory responsibilities when it comes to protecting adults and children from harm. Our team members were trained so they know what action to take if any person discloses an incident or researchers witness an incidence of abuse, violence, exploitation or neglect during the evaluation.

OPM has a safeguarding policy that stipulates overarching principles for working with vulnerable people, including children and young people under the age of 18. All staff, subcontractors and anyone working on behalf of OPM are required to adhere to this policy.

OPM places a mandatory obligation on all staff and contractors to immediately report any concerns, suspicions, allegations and incidents that indicate actual or potential abuse of vulnerable people. For this fieldwork, OPM developed a specific Safeguarding Framework which established a Safeguarding Committee and a child protection reporting protocol. All fieldworkers going to schools were trained in the Framework. More details are available on request.

## E.3 Anonymity

All data collected by the evaluation is anonymised. The quantitative data sets have been made publicly available but anonymised. The research team was trained in confidentiality and signed agreements to keep the responses confidential. In the qualitative report, all schools, communities and districts were anonymised and given a number instead of their name. Where identifying facts were included in quotations, these were removed.

## E.4 Paris Declaration

According to the OECD,<sup>71</sup> the Paris Declaration (2005) is “a practical, action-oriented roadmap to improve the quality of aid and its impact on development. It gives a series of specific implementation measures and establishes a monitoring system to assess progress and ensure that donors and recipients hold each other accountable for their commitments.” This evaluation has been implemented in accordance with the principles of the Paris Declaration as follows:

- **Ownership.** The evaluation has sought country ownership through the establishment of an impact evaluation Reference Group, chaired by senior members of the relevant government ministries, and with membership from technical officers in government as well as national non-governmental representatives.
- **Alignment.** The EQUIP-T programme is aligned with Government of Tanzania objectives, and has made substantial use of government systems (including instructional material development,

<sup>71</sup> See <https://www.oecd.org/dac/effectiveness/parisdeclarationandaccraagendaforaction.htm>

teacher training institutions, public financial management systems – but more broadly it was all implemented using the government education system).

- **Harmonisation.** The development partners coordinate through the education development partners group, in which they seek to harmonise their support to government and minimise duplication and gaps. In the case of EQUIP-T, the regions selected for EQUIP-T intervention were selected so as to avoid the regions benefiting from other donor programmes. The EQUIP-T MA has worked with these other programmes to share materials and learning. The evaluation team sought to capture information from the other donor-supported programmes in order to understand the implications of these other programmes for the evaluation results. The wider development partners group is also represented on the Reference Group.
- **Results.** The evaluation is explicitly focused on measuring the results of EQUIP-T at the outcome and impact level.
- **Mutual accountability.** The evaluation findings are intended to provide accountability, both of the implementing partners to DFID, and of DFID to the government of the UK and Tanzania. The presence of the Government of Tanzania on the Reference Group provides opportunity for the Government of Tanzania to be held accountable for its part in the success of the programme.

Furthermore, the evaluation sought to increase local capacity through working with a number of national team members in the planning, data collection, analysis and dissemination.

## Annex F Summary of PTP objectives and findings from fieldwork on PTP activities

This note was developed to provide some explanation of why there are very differing stakeholder views on the success of PTPs, as revealed by consultations on an earlier draft of the final endline report. A background document review carried out to inform this note is available on request.

### F.1 The objectives of PTPs

In order to evaluate whether EQUIP-T has been successful in achieving its intended objectives for PTPs, these objectives need to be clear. In this regard, there are many interpretations of what PTPs are supposed to do (activities and outputs) and what they should thereby achieve (outcomes). This ambiguity can help explain the different views of stakeholders on whether the intervention has been successful. The question is, successful at what? This note sets out the outcomes and activities identified as objectives of PTPs in programme documentation and interviews with MA staff.

#### Outcomes

PTPs are an intervention under component 4 which should therefore contribute to the component's overall outcome objective: "strengthened community participation and accountability." EQUIP-T documents refer to the purpose of PTPs being "to increase parents' representation and bring them closer to the classroom in order to develop stronger home-school partnerships." The more detailed objectives frequently referenced by programme documentation relate to the PTP enhancing links between the school and community, with specific aims around:

- increasing information flow/sharing in both directions (for parents to understand what happens in schools, and teachers to be more aware of community background and cultural norms), and
- empowering parents and the wider community through a sense of ownership and role in holding the school to account.

There are a number of references to the "wider community", implying the 14 parent PTP representatives are not meant to be the only community members /parents engaged by the PTP. In fact, the initial training on establishing PTPs explains that "All parents of children [at] this school are automatically members of the PTP as are all teachers currently working at the school." And goes on to explain that the core 21 members are representatives who will "be given special assignments and roles in support of the school." For ease, this evaluation refers to the PTP members meaning the 21 core representatives only.

Documents on the purpose of PTPs also refer to facilitating improved classroom teaching. This is to be achieved by a direct presence of PTP members in the classroom, having oversight of the classroom, teachers and performance, but also presumably indirectly through the community involvement and increased accountability set out above. (The detail of this mechanism is not explained in documentation.)

Having outcomes described using an array of general terms such as ownership, empowerment, and accountability, leaves the meaning of success at this level open to wide interpretation.

## Activities

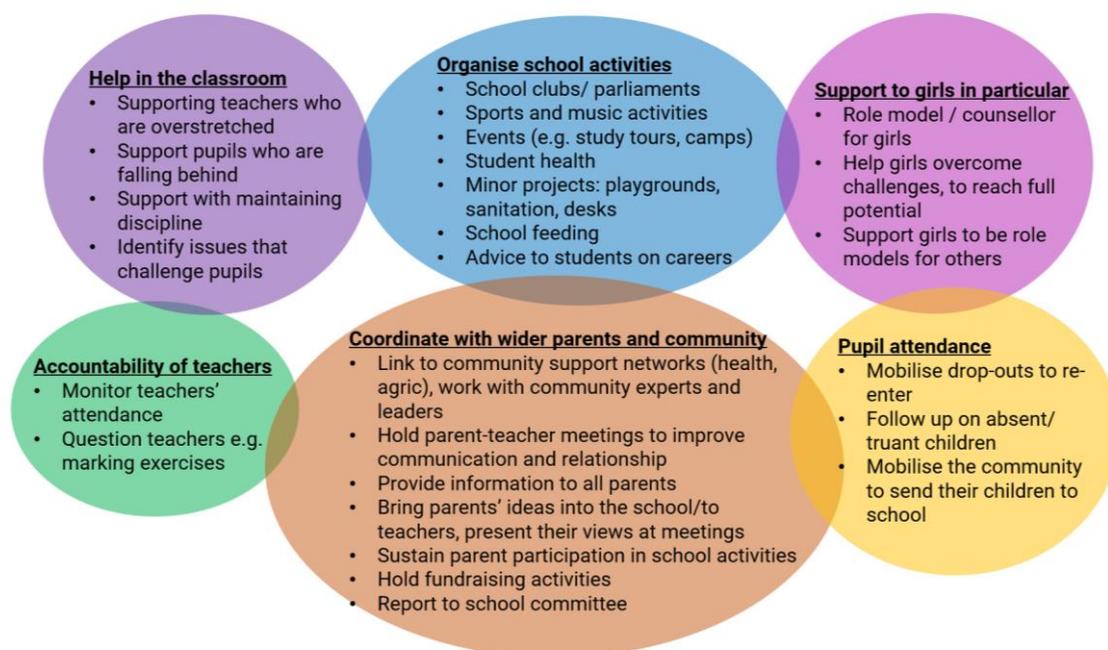
Whilst the activities of PTPs are not directly pre-determined – they are “meant to be decided at school level based on each school’s needs and priorities” – programme documentation does include a number of suggested activities. These suggestions are included in training for school committees on PTP establishment and things PTPs can do which do not require the grants. (Their role in relation to the grants is not discussed here.)

The suggested activities for PTPs described in the documents fall into six broad categories, though there are many overlaps, and other ways of grouping the activities could be envisaged. The activities are listed in Figure 17 below, and the categories are:

- Help in the classroom
- Organise school activities
- Support to girls in particular
- Follow up on pupil attendance
- Coordinate with wider parents and community
- Follow up on accountability of teachers

These are not all given equal weight – it should be noted that accountability of teachers is rarely mentioned. Meanwhile much more is said about the role of girls’ education in relation to the second PTP grant, but the activities set out in Figure 17 are focused on things PTPs can do without grants.

**Figure 17. Suggested activities for PTPs (without needing PTP grants)**



What this does show is that there are a wide range of supposed activities and roles for PTPs, which can explain why PTPs themselves are not clear what they should do (it would depend heavily on the leadership and quality of the training passed on by the head teacher and WEO). It also reveals why stakeholders have differing views on the success of PTPs – they may be delivering some, but not all of these activities.

Finally, it is not obvious from the documentation how implementing a selection of these activities would lead to the intended outcomes (such as empowerment and accountability), as these are put in very general terms.

## F.2 Response from fieldwork on PTP activities

The table below sets out the evidence found from the endline fieldwork, organised against the categories of activities from Figure 17.

For the quantitative fieldwork, head teachers were asked to name all the types of actions/activities that PTPs had undertaken in the last school year, and these were then recorded against a pre-determined list of activities<sup>72</sup>. In total 68% of PTPs had taken some action. For the qualitative fieldwork, the focus of the research was on what made some schools and parents successful in implementing the PTP initiative and engaging the community overall. Respondents included head teachers, PTP teacher and parent members and community leaders. Questions asked included:

What are the roles and responsibilities of PTP members? Which activities take the most time? What are the most challenging activities for PTPs? What are the PTP’s achievements? What does the PTP do to seek parental engagement? What were the PTP grants spent on? What projects does the PTP have?

The results from the qualitative endline are not intended to be representative of all schools (these were chosen as being schools successful in implementing PTPs, and not all schools gave the same answers). However this shows the range of responses that were received from the six schools sampled.

**Table 17. Mapping of findings of PTP activities against suggested activities**

Suggested activities	Quantitative endline	Qualitative endline – examples from six schools with active PTPs
Help in the classroom	<ul style="list-style-type: none"> <li>• Less than 2% of PTPs had a community member assist in the classroom</li> <li>• Less than 2% of PTPs took action to provide extra tuition/classes</li> <li>• 4% of PTPs took action to provide exams/tests</li> <li>• Less than 2% of PTPs took action to provide extra teaching and learning materials</li> </ul>	<ul style="list-style-type: none"> <li>• Try to find solutions to shortage of teachers (e.g. volunteer teachers)</li> <li>• Making follow up on students’ performance and challenges</li> <li>• Discuss challenges with individuals, and solutions with teachers and the head teacher</li> <li>• Identify children who are behind and organise lessons after-school</li> <li>• Assist with discipline in large classes</li> </ul>
Organise school activities	<ul style="list-style-type: none"> <li>• 2% of PTPs supported extra-curricular activities</li> </ul>	<ul style="list-style-type: none"> <li>• Supervising sports competitions between schools</li> <li>• Organising hostels or camps for Standard 7 pupils</li> <li>• Organising food before vaccinations</li> <li>• Supporting school security and boundaries</li> <li>• School farm plot</li> </ul>

<sup>72</sup> The pre-determined list was developed based on responses given by head teachers during extensive pretesting and piloting of the instruments. There was also an opportunity to give other answers and explain them.

<i>Infrastructure</i>	<ul style="list-style-type: none"> <li>• 10% of PTPs took action to improve infrastructure</li> </ul>	<ul style="list-style-type: none"> <li>• Supporting classroom and toilet construction, minor repairs</li> </ul>
<i>School feeding</i>	<ul style="list-style-type: none"> <li>• 9% of PTPs took action to provide school feeding</li> </ul>	<ul style="list-style-type: none"> <li>• Organise school feeding using contributions from parents</li> </ul>
Support to girls in particular	<ul style="list-style-type: none"> <li>• 6% of PTPs took action which was intended to improve education for girls</li> </ul>	<ul style="list-style-type: none"> <li>• Sensitising girls and parents on avoiding early pregnancy</li> <li>• Sensitising parents on the importance of girls education</li> <li>• Projects to purchase sanitary pads</li> <li>• Sensitising girls and parents on the use of sanitary pads</li> </ul>
Follow up on pupil attendance	<ul style="list-style-type: none"> <li>• 46% of PTPs took action to improve pupil attendance and punctuality</li> </ul>	<ul style="list-style-type: none"> <li>• Sensitise parents on the importance of education and attending school</li> <li>• Mobilise for registration</li> <li>• Visit classrooms to check for absenteeism</li> <li>• Check exercise books for dates and attendance</li> <li>• Follow up on absenteeism</li> <li>• Visit parents of absent children, ask why the child does not attend school</li> <li>• Write to parents of truant children</li> <li>• Speak to Community leader and at village meetings</li> <li>• Bring any truant pupils (and information about them) into school</li> <li>• Ask community members to bring any truant pupils (and information) into school</li> <li>• Send other pupils to find truant children</li> <li>• Call the school committee</li> </ul>
Coordinate with wider parents and community	<ul style="list-style-type: none"> <li>• No response specifically on this</li> </ul>	<ul style="list-style-type: none"> <li>• Using village meetings / community meetings / village government meetings to:</li> <li>• discuss absenteeism / sensitise on the importance of education (attendance, supporting hard work/success in exams)</li> <li>• sensitise parents on the importance of girls' education</li> <li>• ask parents to come to school meetings</li> <li>• Visit individual parents on the above</li> <li>• Visit the Community Leader</li> <li>• Send messages to parents via children or the community leader.</li> <li>• Provide information to parents, tell parents about challenges in classrooms, about graduation</li> <li>• Discuss class problems with the teacher</li> <li>• Resolve school boundary issues between school and community</li> <li>• Present information to the school committee</li> </ul>
<i>Fundraising</i>	<ul style="list-style-type: none"> <li>• 6% of PTPs took action related to IGAs</li> </ul>	<ul style="list-style-type: none"> <li>• Establishing small IGAs to raise funds</li> </ul>

	<ul style="list-style-type: none"> <li>• Less than 1% held fundraising activities</li> </ul>	<ul style="list-style-type: none"> <li>• Farming small plots (for fundraising or for school feeding)</li> <li>• Visit households to request contributions for classroom construction, making bricks, desks, school feeding, examination camps, volunteer teachers.</li> </ul>
Follow up on accountability of teachers	<ul style="list-style-type: none"> <li>• 8% of PTPs took action to improve teacher attendance and punctuality</li> </ul>	<ul style="list-style-type: none"> <li>• Follow up if teachers are teaching</li> <li>• Ask pupils what they have learnt</li> <li>• Check exercise books</li> </ul>

The findings suggest the following:

- Help in the classroom – successful PTPs are entering classes, although they do not help directly with teaching, they help to identify individual and class-wide challenges and solutions.
- Organise school activities – a sizeable minority of PTPs are involved in either infrastructure provision or organising school feeding. Other activities in successful PTPs include organising examination camps and sports competitions.
- Support to girls in particular – successful PTPs are involved in sensitisation around girls' issues – sensitising girls themselves and their parents on using sanitary pads and avoiding pregnancy. Many PTPs bought sanitary pads (often linked to the second PTP grant).
- Follow up on pupil attendance – this is the most common activity of PTPs, and most active PTPs are taking actions in this area. PTPs check for absenteeism by checking exercise books and visiting classrooms. Actions include speaking and writing to parents of absent children, organising for absent children to be brought into school, and using other platforms (village leaders and committees) to share information about the importance of attending school. PTP members help to mobilise parents to register their children to enter school.
- Coordinate with wider parents and community – successful PTPs have been engaged with community meetings to sensitise around the importance of education and parental engagement with the school. PTPs provide information to parents. PTPs help resolve boundary issues between the school and community. A major part of PTPs engaging with parents is to raise resources and contributions for the school, either financial or in-kind.
- Follow up on accountability of teachers – a small minority of PTPs follow up with teachers, and examples include asking pupils what they have learnt to assess teacher progress and attendance.

## Summary

Successful PTPs identified in the evaluation's fieldwork are active and carrying out a number of the suggested activities, but this is not typical of PTPs in general. The intended outcomes of PTPs are open to different interpretations because of the general terms used and lack of concrete description of what success would look like in schools-communities at this level. The pathways between implementing a selection of suggested activities and achieving the intended outcomes are far from clear. The link to improved information flow and participation in school improvement may be obvious, but it is not apparent that these activities would lead to the PTP being able to hold the school to account, and shift the power balance between schools and communities. It is thus not surprising that stakeholders have different perspectives on success, as they may be judging the success of PTPs at different levels or with different interpretations of what the outcomes actually mean.

### F.3 Source documents

The following source documents were reviewed to establish the intended outcomes and activities of PTPs. The background document is available upon request.

#### Documents

- EQUIP-T (2015a) Ruzuku ya Ushirikiano wa Wazazi na Walimu. Kiongozi Cha Mwezeshaji. (PTP grant Trainers Guide, Final version)
- EQUIP-T MA (2015b) 'EQUIP-Tanzania Annual Report 2015'. United Republic of Tanzania (URT), PMO-RALG, MOEVT, and DFID.
- EQUIP-T MA (2015c) Summary of Component 4; Annex 89 to EQUIP-Tanzania Annual Report.
- EQUIP-T MA (2016) 'EQUIP Tanzania, 2016 Annual Report External Version'. URT, DFID.
- EQUIP-T MA (2017a) EQUIP-Tanzania Proposed Programme Extension. (Extension proposal as at 17<sup>th</sup> February 2017)
- EQUIP-T MA (2017b) 'EQUIP Tanzania, 2016-17 Annual Summary External Version'. URT, DFID.
- EQUIP-T MA (2018) Report on Parent Teacher Partnership Grant Implementation. Key Document 15 in 2018 Annual Report.
- EQUIP-T MA (no date, 2019 or 2020) Strengthening Parental Engagement in Schools: Parent Teacher Partnerships. EQUIP-Tanzania Practice Paper.
- IPSOS (2015) EQUIP-T Annual Monitoring Survey. Draft Final Report. Annex 61 of EQUIP-T MA 2015 Annual Report.
- OPM (2019a) EQUIP-Tanzania Impact Evaluation. Endline Quantitative Technical Report, Volume I. Results and Discussion.
- OPM (2019b) EQUIP-Tanzania Impact Evaluation. Endline Planning Report: Part II. Qualitative research, cost study, and endline synthesis.
- PO-RALG and MoEVT (2015) School Committee Training Guide. Education Quality Improvement Programme in Tanzania. Funded by DFID.
- PO-RALG (2019) RE: Sustainability of EQUIP-Tanzania Best Practices. Letter from PO-RALG to EQUIP-T National Coordinator, cc. DFID. 16<sup>th</sup> January, 2019.

#### Interview notes

- OPM Interview with component leads, January 2016
- OPM Interview with component leads, January 2018
- OPM and MA TOC workshop, January 2018
- OPM Interview with component lead, February 2019

## Annex G Recommendations for government: details

This annex sets out the recommendations for government from this report, but provides additional details. First, it includes a list of which institution (ministry, department and so forth) is responsible for implementing the recommendation. Second, it categorises the recommendations by priority, as primary or secondary recommendations. Primary recommendations are those deemed to have large impact, even if at contained scale, or medium impact at national scale; and medium or low additional costs. Secondary recommendations are those with lower size or scale of impact or very high additional costs. Whilst all recommendations are made because the evaluation team feel they are important and beneficial, they are categorised in order to aid decision makers. Finally, a double asterisk (\*\*) is included to indicate recommendations that build on commitments and efforts already made and would be more straightforward to get started.

**Table 18. Recommendations with responsible institutions and priority**

#	Recommendation	Institutions responsible	Priority
1	<p><b>Under the NF-TCPD, continue the in-service training model in EQUIP-T regions and scale this up to other regions that do not already have an effective school-based model.</b> The EQUIP-T in-service training model has been effective in improving teacher confidence and morale, and in empowering teachers with a range of approaches which have contributed to improved learning outcomes. The model conforms to the principles and modalities set out in the NF-TCPD, and the numeracy, literacy, and other modules have been approved by TIE.<sup>73</sup></p> <p>The model should continue at school level, which is lower cost, effective at providing ongoing support, and means teachers receive training despite high turnover. However, there does need to be intermittent face-time with experts (such as TTC tutors or TRC staff) to refresh key ideas and troubleshoot when there are problems. The revamping of TRCs, under the NF-TCPD, will help to support this need.</p> <p>The school-based model should be institutionalised through the formal recognition of the in-service coordinator role in schools (which may need promotion and additional salary), and by building this into the monitoring and quality assurance responsibilities of WEOs and SQAs, as laid out in the NF-TCPD.</p> <p>The participation and completion of in-service training should be built into the teacher career progression framework to incentivise taking part. The NF-TCPD states the intention to recognise and certify TCPD based on participation in standard modules that are accessible to all teachers.<sup>74</sup></p> <p>The best practices from the in-service training content should be integrated into pre-service training programmes.</p> <p>At the time of this evaluation's fieldwork, it was clear that teachers faced challenges in managing large classes. The 'general effective pedagogy' module, rolled out by EQUIP-T after the qualitative fieldwork, will have been important in providing content on methods for large classes.</p> <p>The government will need to increase budget allocation to education,</p>	TIE, PO-RALG (Pre-primary and primary education section) and MoEST (Teacher training division).	Primary **

<sup>73</sup> The NF-CPD (p. 9) states that TCPD should be grounded in collaborative, inclusive, gender-responsive, and participatory learning. It also emphasises the importance of non-residential models which allow teachers to be on task during school hours (p. 22). These elements are all central to EQUIP-T's model.

<sup>74</sup> See MoEST (2019).

	including through LGAs, to sustain the in-service training model.		
2	<b>Make use of the TLMs distributed under EQUIP-T.</b> The reading books are not currently being used, and they are a valuable resource. The government could include lessons on how to use reading books into both pre-service and in-service training, and include use of the books in curriculum guidelines (such as a minimum number of minutes per week for pupil reading practice).	TIE and MoEST (SQA division)	Secondary **
3	<b>Review, with partners, the support for teachers on teaching children who do not speak Kiswahili as their mother tongue.</b> Many teachers are in contexts where many pupils do not speak Kiswahili as their mother tongue. This evaluation shows that these pupils are far behind their classmates in Kiswahili and maths, and wider evidence shows it could take many years for these pupils to catch up (Collier, 1989). Teachers would benefit from support in acquiring skills in approaches to teaching pupils in multilingual classrooms, and in putting these into practice. Such techniques could include games, group work, and use of translanguaging <sup>75</sup> (Heugh et al., 2019). Interventions to help children catch-up before starting, such as the SRP, may also be beneficial, and this would be a useful focus for future research in Tanzania.	TIE, PO-RALG (Pre-primary and primary education section) and MoEST (Teacher training division).	Secondary
4	<b>Continue the focus on reducing classroom shortages and recruiting more teachers.</b> The evaluation confirms the extremely large pupil-to-classroom and pupil-to-teacher ratios and the challenges faced by teachers in this context. Building more infrastructure and recruiting more teachers are necessary to alleviate these barriers to effective teaching.	PO-RALG (Pre-primary and primary education section)	Secondary
5	<b>Review the teacher management policies which lead to high absenteeism and turnover.</b> The extremely high level of teacher absenteeism and turnover affects the usefulness of any intervention to improve teaching quality in the classroom. Improved monitoring and accountability for classroom attendance, and reducing the rate of transfers, would help address this. The recommendation to review transfer and turnover is also relevant to head teachers, WEOs, and LGA officers.	ADEM and PO-RALG (Pre-primary and primary education section)	Primary
6	<b>If the government is planning to continue using SIS, it needs improvements from the version seen by the evaluation team.</b> It should be reviewed in terms of ease of data entry for school staff and how to make the data useful at the school level. The SIS should not replace existing data collection systems until it is proven to work and all glitches have been removed for a whole annual cycle. More than one person per school should be trained in how to enter and use the data.	PO-RALG (EMIS - unit)	Secondary **
7	<b>Continue and scale-up DEMs across the country.</b> The introduction of DEMs has been successful at improving relationships between WEOs and the education department, sharing experience and learning, and allowing more efficient information sharing and management of schools. Ideally, DEMs should include an aspect of demand-driven training (akin to WEO CPD). Based on the findings of the evaluation, this roll-out will need to involve some training, rather than just self-reading materials, with examples of best practice, and continued leadership and emphasis from all levels of government to maintain and institutionalise the practices.	ADEM and PO-RALG (Pre-primary and primary education section)	Secondary **
8	<b>Produce a standardised manual on WEOs' roles and roll these out through DEMs.</b> WEOs lack previous training on their roles and	ADEM and PO-RALG	Secondary **

<sup>75</sup> Translanguaging includes a range of processes in which bi-/multilingual people make use of the knowledge they have of many languages and how to use these languages. This can include alternating between two or more languages.

	responsibilities, and appreciate the training they have received under EQUIP-T; this suggests the gap is likely to apply in the rest of the country and that a manual would be useful for new WEOs at least. After initial roll-out, the manual could be used within ongoing DEMs to discuss how WEO responsibilities should be performed and to guide the CPD aspect of these meetings.	(Pre-primary and primary education section)	
9	<b>Continue providing responsibility allowances for WEOs and strengthen accountability for their performance.</b> The experience of the WEO grant under EQUIP-T has been that it unlocked more frequent visits to schools by WEOs. The government introduced a responsibility allowance in 2016, and although WEOs in EQUIP-T regions will now lose the value of the EQUIP-T grant, they will still have more resources than in the past, which is likely to make a substantial contribution to WEOs' performance. Although the use of the responsibility allowance is not stipulated (though perhaps should be), the government should enforce performance management of WEOs so that they are accountable for making visits to schools and supporting school improvement.	PO-RALG (Pre-primary and primary education section)	Secondary **
10	<b>Review how to strengthen parental engagement and school accountability to the community, drawing on the best practices of SCs and PTPs.</b> SCs play an important and recognised role, including in engagement with parents. The SC parent member is seen as a preferred channel for parents to raise concerns, rather than via the PTP. In many cases, introducing PTPs has caused confusion and minimal additional benefit in terms of wider parental engagement and school accountability. However, some PTPs are carrying out activities which are seen as beneficial within the school, even if they are not recognised as bringing the wider parental body closer to the school. It would be worthwhile considering options for encouraging the continuation and scale-up of these types of parent-led activities – involving PTPs and/or SCs. Future efforts to strengthen wider parental engagement and school accountability should recognise the central role of the SC, as well taking into account the constraints to participation and empowerment faced by parents, in deciding how to engage and represent them.	ADEM and PO-RALG (Local government division and Education administration division)	Secondary
11	<b>Consider continuing distribution of school grants for IGA or school improvement.</b> The PTP grants and IGA grant have energised community mobilisation and participation, in terms of labour, in-kind resources, and interest in the school. The PTP grants allowed direct school improvement and gender welfare activities, whereas IGA projects, if successful, create an ongoing income stream for the school. The PTP grant does not require a PTP to be successful – the same aim could be achieved through the SC – and thus this recommendation can be considered separately. The cost analysis shows that the cost of these grants is actually rather small in comparison with current LGA budgets. Distributing grants for IGA or school improvement – in new regions which have not received them already – would help the SC to engage more with the school and the wider community.	PO-RALG (Education administration division) and MoEST (Policy development division)	Secondary
12	<b>Embed any new practices throughout the management chain.</b> Where activities are continued or scaled up to other regions, they need to be embedded in the responsibilities and monitoring systems throughout the chain – in how head teachers supervise teachers, how WEOs and SQAs supervise and help ensure quality in schools, how LGAs monitor WEOs, right up to regions and central ministries. At LGA level, DEMs are a platform to collect information on practices from the school and WEO level (such as in-service training). Regions should continue collecting performance and activity reports from LGAs as this accountability has	ADEM and PO-RALG (Pre-primary and primary education section) and MoEST (Policy development division and	Primary

	contributed to an increased focus on results.	SQA division)	
13	<p>Recent trends in the education sector budget suggest that paying for the three activities studied here – teacher in-service training, PTP grants, and IGA grants with initial training – will not be affordable for the LGAs without additional financial support. The central government should take two actions if it wishes to continue replicating and scaling-up these activities across the country.</p> <ul style="list-style-type: none"> <li>• <b>Recognise the cost burden and provide sufficient budget for these activities</b>, whether that budget is held and spent at national, regional, LGA, or school level. Government should not expect LGAs and schools to implement these activities if no provision is made for the costs. Government should also consider ways to reduce costs whilst maintaining satisfactory quality, as reviewed in the cost study (OPM 2020b).</li> <li>• <b>MoEST and PO-RALG should strengthen the case for additional spending in the education sector</b>, to put to MoFP and Parliament. This requires reviewing and assessing the evidence, and communicating this evidence to MoFP, the Cabinet, and parliamentarians, who each have a role in approving the final budget. Furthermore, MoFP should present annual budget allocation and disbursements at the joint annual sector review to provide additional accountability.</li> </ul>	PO-RALG (Local government division), MoEST (Policy development division) and MoFP (Government budget division and Planning division)	Primary

## About the project

The independent Impact Evaluation of the Education Quality Improvement Programme in Tanzania (EQUIP-T) is a six-year study funded by the United Kingdom Department for International Development (DFID). It is designed to: i) generate evidence on the impact of EQUIP-T on primary pupil learning outcomes, including any differential impacts for girls and boys; ii) examine perceptions of effectiveness of different EQUIP-T components; iii) provide evidence on the fiscal affordability of scaling up EQUIP-T post-2019; and iv) communicate evidence generated by the impact evaluation to policy-makers and key education stakeholders.

EQUIP-T is a Government of Tanzania programme, funded by UK DFID, which seeks to improve the quality of primary education, especially for girls, in nine regions of Tanzania. It focuses on strengthening professional capacity and performance of teachers, school leadership and management, systems which support district management of education, and community participation in education.

