

Bristol Working Papers
in Education Series

 University of
BRISTOL
School of Education

Working
Paper
#06/2018

Scan of International Approaches to Teacher Assessment

Artemio Arturo Cortez Ochoa
Professor Sally Thomas
Professor Leon Tikly
Helen Doyle

School of Education
University of Bristol
35 Berkeley Square
Bristol BS8 1JA

bristol.ac.uk/education/research/publications



Scan of International Approaches to Teacher Assessment

**Artemio Arturo Cortez Ochoa, Professor Sally Thomas,
Professor Leon Tikly and Helen Doyle, University of Bristol**

April 2018

Funder: Mastercard Foundation

Contact: s.thomas@bristol.ac.uk

This online paper may be cited or briefly quoted in line with the usual academic conventions, and for personal use. However, this paper must not be published elsewhere (such as mailing lists, bulletin boards etc.) without the author's explicit permission, or be used for commercial purposes or gain in any way.



SCHOOL OF EDUCATION

Contents

Executive Summary.....	3
Introduction – terms of reference	7
Methodology.....	7
Background	8
Challenges of secondary education in sub-Saharan Africa (SSA).....	8
The Secondary Education Teachers Initiative (SETI).....	8
Context – Rwanda.....	9
Lessons from elsewhere in sub-Saharan Africa	12
Literature review.....	16
Why teacher quality?.....	16
International literature	18
Previous literature reviews relating to Teaching Evaluation Systems.....	18
Summary of findings from previous literature reviews	22
Literature relating to Teaching Evaluation System goals and purposes.....	23
An essential conceptual differentiation.....	24
‘Standards’ and ‘Frameworks’ for Teaching	25
Literature relating to Standards for Teaching.....	25
An overview of existing Frameworks for Teaching.....	29
Summary of findings from existing standards and frameworks.....	35
Subject-specific frameworks	36
Instruments currently used for Teaching Evaluation Systems	39
Classroom observations.....	40
Teacher interview/dialogue/appraisal with colleagues and school authorities.....	43
Teacher testing	44
Peer assessment	44
Teacher portfolios.....	45
Classroom artefacts	46
Student surveys.....	46
Teacher self-assessment.....	47
Value-added student outcomes.....	47
Concluding remarks relating to instrument selection and design.....	48
Recommendations	49

References 52
Appendices..... 60

Executive Summary

The research on which this report is based was commissioned by the MasterCard Foundation (MCF).

The research was conducted between August and September 2017.

The aim of the research, as set out in the MCF Terms of Reference, is to provide an overview of international approaches, best practice and instruments used in the development of Teaching Evaluation Systems (TES), which is then used to formulate recommendations for the development of such a system for use by MCF to evaluate the Secondary Education Teachers Initiative (SETI) it is implementing, initially in Rwanda but which can be adapted to other country contexts.

Methodology

The research took the form of a focussed literature review, with priority being given to the most recent research evidence and cases of TES from sub-Saharan Africa (SSA) and other low-income contexts.

The recommendations were formulated through team discussion of the research findings.

Summary of findings

Context:

- The SETI coordinates activities of partner organisations dedicated to improving the quality of teaching and learning in secondary schools, and the establishment of Centres for Innovative Teaching and Learning which will target key themes in secondary teaching and learning across the SSA context.
- Currently the initiative is being set up in Rwanda, but will be extended to other countries in the future.
- The TES, then, is required to measure the impact of the SETI on the quality of secondary teaching, initially in Rwanda, but with the ability to be used in other contexts in SSA.
- The main issues relating to the quality of teaching in secondary schools in Rwanda are the high proportion of un- and under-qualified teachers (especially in rural areas), the introduction of a competence-based curriculum (the delivery of which teachers have not been adequately prepared for), and poor morale among teachers with problems relating to recruitment and retention.
- Features of the Rwandan context which are advantageous to the SETI are the ongoing development of a set of teacher standards by the government, existing incentives to attract and retain high quality teachers, and the existence of a School-Based Mentoring Programme which could be incorporated into a TES.

Lessons from elsewhere in SSA

Although most research on TES has been conducted in highly industrialised and high-income contexts, there are a small number of case studies from elsewhere in SSA which are instructive when contemplating designing a TES to evaluate the SETI. Research has established that the quality of teaching is as important for student outcomes in low income countries as it is in high income countries. Case studies examined for this research were from Kenya, South Africa, Malawi, Zambia, Tanzania and Botswana, and revealed the following issues:

- There is a need for high quality training for evaluators in the use of the instruments being used for evaluation (Kenya).
- All stakeholders need to be consulted about the TES from the very beginning if they are to be fully supportive of it (South Africa).
- Any TES should be supported by a set of standards which lay out the expectations of teachers (Malawi).
- For a TES to contribute to an improvement in teaching, it needs to include a strong element of CPD (Zambia, Tanzania).
- To ensure the cooperation of all teachers, a TES needs to be implemented equitably across all subject areas and levels of experience (Zambia).
- TESs need to be revisited frequently to ensure that all parties are cognisant of their aims and objectives (Botswana).
- The World Bank Service Development Indicators recently used in Kenya and Uganda have been developed for benchmarking standards in primary education, and thus have limited relevance for the SETI.

Lessons from the literature review

Research is providing more and more evidence that student outcomes are closely related to teacher behaviours, and this is especially the case for more deprived students. The current focus is on research that establishes the particular teacher behaviours which lead to students making the most progress. However, this research is being conducted in high-income countries, and does not take account of the more challenging contexts of SSA (such as large class sizes and scarce resources). Thus, while lessons can be learned about general principles for the design and implementation of a TES, contextualising the system is vital.

The literature review examines findings from five of the most recent reviews of TESs from a range of scales and locations. It finds that the most effective TESs have the following features: involvement of

all stakeholders during formulation of the system, clarity of purpose, guidance from standards or frameworks to establish the characteristics of good teaching, gathering of evidence from a range of sources, a link to identifying areas for improvement and possible CPD for teachers, clear planning and governance, an appreciation of the context in which the TES will be used, and acknowledgement of both positive and negative outcomes of its implementation.

Some of the most widely used standards and frameworks are then reviewed, and the advantages and disadvantages of each one for a context like Rwanda is summarised. It is noted that standards and frameworks constitute the basis for what is considered to be good teaching in the context in which they are applied, and are then the foundation for choosing the instruments which will be used to evaluate teaching quality. They should also be designed to facilitate comparisons of teaching quality both at one point in time and over a period of time.

Instruments used for TES

The instruments that are most commonly used in TESs are evaluated, bearing in mind the particular context of Rwanda and SSA. These are: classroom observations, teacher interviews, teacher testing, peer assessment, teacher portfolios, classroom artefacts, student surveys, teacher self-assessment and value-added student outcomes. The critical considerations when selecting instruments are found to be: the resources that are available to the assessor, the purpose of the evaluation (e.g. formative, summative or both), and the context in which the instruments will be used. It is also clear that evidence is needed from a range of sources, one of which should be classroom observation.

Recommendations

Over-arching principles for the establishment of a TES:

- Explicit governance which includes inputs from all stakeholders and collaborative design.
- Alignment with the context and culture in which it is being used.
- Evidence-based.
- Part of an integrated system which has an overall purpose of teacher and therefore school improvement.
- Alignment with existing and planned policies.
- Careful consideration of the tensions between the formative and summative elements of the systems, especially in contexts where there is low morale and poor recruitment and retention of teachers.

Recommendations for each stage of establishment:

Stage 1 – establishing teaching standards

- Standards need to be appropriate to the context, including the constraints under which teachers work – they should be realistically achievable by the majority of teaching staff.

Stage 2 – developing a framework

- The framework needs to be based on the teaching standards.
- It needs to be applicable to teachers at any and all stages of their career.
- It should provide for identification of CPD needs.
- It may specify particularly appropriate instruments.

Stage 3 – developing instruments for gathering evidence and making judgements

- Evidence to support judgements should be collected using a range of instruments, prioritising those which enable the identification of CPD needs.
- When selecting instruments, account must be taken of the context in which they will be used, for example levels of training and availability and reliability of data.

Stage 4 – implementation of the TES

- Training is crucial, both for evaluators and teachers who will be evaluated.
- Provision must be made for resources for any CPD that may arise from evaluations.
- Implementation must be considerate of teacher workload.
- Implementation must be equitable across the whole teaching body.
- Aims and objectives of a TES need to be revisited regularly to ensure that it is fit-for-purpose and all teachers are fully cognisant.
- As far as possible, a TES should make use of any existing systems and policies related to teacher evaluation.
- Care must be taken to address any unintended consequences of the TES as soon as possible.

Introduction – terms of reference

In accordance with the MCF Terms of Reference, this report will provide an overview of international approaches, best practices and methods/instruments for assessing teacher competence and performance, referred to as Teaching Evaluation Systems (TES). This overview will then be used to formulate recommendations for appropriate approaches and instruments to be used by the MasterCard Foundation (MCF) in evaluating the Secondary Education Teachers Initiative (SETI), initially in Rwanda but which can be adapted to other specific country contexts.

The report begins with an examination of the context in which MCF is implementing the SETI. Different approaches to TES from across sub-Saharan Africa and issues arising from these approaches are explored. There is then a review of the most recent literature relating to TESs, which is used to establish the key principles that should underpin such systems. The full range of instruments available for assessment of teaching quality is considered, with specific reference to their suitability for the SSA context. Finally, all of the findings from the previous sections are brought together to formulate recommendations for the establishment of a TES for the MCF SETI project.

Methodology

Following preliminary research and clarification of terms of reference with MCF, a focussed literature review was completed. Initial search terms included ‘teacher quality’, ‘teaching quality’, ‘teacher assessment’, ‘teacher evaluation’ and ‘assessment of teacher competence and performance’. Papers identified using these terms were then scrutinised for further relevant material. Particular attention was paid to the most recently published material, and research relating to the context in which MCF is working (i.e. lower income countries and sub-Saharan Africa).

The relevant material was studied and organised under themes, as follows:

- Analysis of the Rwandan context.
- Case studies of TESs from SSA.
- Previous literature reviews relating to TESs.
- International examples of TESs, including examples of Standards and of Frameworks.
- Discussion of different instruments used in operationalising TESs.

Each theme was written up individually, although there is inevitably some crossover between themes. Repetition has been avoided as far as possible.

Finally, discussion between the team members facilitated the formulation of recommended approaches and instruments for teacher assessment that can be used by MCF for the Secondary Education Teachers Initiative.

Background

Challenges of secondary education in sub-Saharan Africa (SSA)

The MasterCard Foundation (MCF) recognises that there is an urgent need to increase the quality and relevance of secondary education in SSA, with research revealing that ‘only a fraction of secondary-aged students...complete secondary education’, and ‘learning outcomes are significantly lower than international benchmarks and many graduating students lack sufficient literacy and numeracy skills’ (MCF 2017 p5). High quality secondary education is essential for poverty alleviation, as it targets people at a stage in their lives when their choices will have ‘consequences for themselves, their families, their communities and societies’ (ibid. p5), and potentially gives young people the relevant skills and knowledge to ‘thrive at work and as citizens’ (ibid. p5). However, there is a lack of funding from other donors for this phase of education. MCF aims to plug this gap, specifically by focussing on the quality of teaching through the Secondary Education Teachers Initiative. Details of this initiative are given below.

The Secondary Education Teachers Initiative (SETI)

The objectives of the SETI are:

1. To attract and retain teachers, especially in remote rural regions.
2. To deliver 21st century curriculum and pedagogy to secondary school students.
3. To improve teacher motivation and elevate the status of the profession.
4. To create enabling and supportive environments for effective teacher practice.
5. To support education leaders to formulate evidence-based policies and plans and foster a network of education innovators. (ibid. p7).

There are two components to the SETI. These are:

- a country level approach, where MCF integrates, facilitates and directs initiatives implemented by partner organisations to improve the quality of teaching and learning, and
- the establishment of up to three Centres for Innovative Teaching and Learning, each of which will target a key theme in secondary teaching and learning with particular relevance to the SSA context.

Currently, MCF is running a country level programme in Rwanda, where it is also developing a Centre to ‘spark innovation and promising practices in the use of ICT for teaching and learning in secondary education across Africa’ (ibid. p8). Within the country level programme, there is a particular focus on training for teachers of Science and Mathematics, as there are problems with teacher shortages and retention rates in these areas. The partner organisations that MCF are working with include the University of Rwanda College of Education, VVOB Rwanda and the Forum for African Women Educationalists (FAWE) Rwanda.

The key principles guiding the SETI are:

- A focus on equity and inclusion.
- A grounding in national priorities and local challenges and opportunities.
- A focus on innovation.
- A focus on systems-level impact.
- Elevation of student and teacher voice, in order to ensure that interventions reflect their needs and to give them a sense of ownership of the initiatives.

The SETI is designed to have a systems-level impact which will lead to significant improvement in the quality of teaching and learning in classrooms throughout Rwanda. MCF requires an ‘effective way to measure improvements in content knowledge (particularly in science and mathematics), the effective use of Information and Communication Technologies (ICT) in the classroom and improvements in pedagogy and instructional practice’ (MCF Terms of Reference). This report will recommend approaches and instruments that will be useful in the endeavour to design a suitable Teaching Evaluation System (TES), while adhering to the guiding principles laid out above.

Context – Rwanda

As stated above, one of the guiding principles of the SETI is that it must be closely aligned with the priorities of the government of the country in which it is being implemented, and take account of local challenges and opportunities. This section of the report considers these factors within the Rwandan context, to ensure that any approaches and instruments are appropriate and in line with this principle.

Government policy regarding secondary education

The Rwandan education system is governed by a complex interaction between a number of government ministries and other agencies. The most significant agency in the governance of the secondary sector, and with particular relevance to this report, is the Rwandan Education Board (REB), established in 2011 to coordinate efforts to increase the quality of education. The REB has

responsibility for curriculum development, teacher development and evaluation, monitoring of education quality and setting of standards (in addition to other responsibilities not relevant to this report). A number of these responsibilities are discharged by Regional Inspectors (RIs) who interact directly with school and local government officials, especially District Education Officers (DEOs) and Sector Education Officers (SEOs). RIs arrange, conduct, report and follow up on school inspections, which include lesson observations, using a Lesson Evaluation Sheet produced by the REB (see Appendix A). Reports are expected to include identification of strengths and areas for improvement, which schools then build in to their improvement plans. However, there does not appear to be any evaluation of individual teachers.

In terms of policy, provision was made for the implementation of a core teaching values and competences profile, for use in the assessment of teachers, as far back as 2007 (Republic of Rwanda, 2007). However, any standards that have been formulated remain in draft form and unavailable to the public.

The two most significant policies which have had an impact on the quality of teaching and learning in recent years are the introduction of a new competence-based curriculum and the establishment of a School-Based Mentoring Programme (SBMP). These policies are described in more detail below.

Competence-based curriculum

The phased introduction of a new competence-based curriculum from 2016 onwards (Republic of Rwanda 2015). The key principles of the new framework are:

- the curriculum and teaching are learner-centred,
- teaching and learning are based on key competences rather than knowledge,
- schools must be inclusive, flexible, transparent and accountable,
- ICT is integrated across all subjects,
- there are cross-cutting issues that make connections between different subject areas.

This new curriculum necessitates a shift from traditional methods of instruction to a learner-centred approach. This is especially significant given that most Teacher Training Institutions in Sub Saharan Africa train teachers in teacher-centred approaches (Doyle et al 2017, José Passos 2009). In recognition of this significant change in pedagogy, teachers were supposed to receive ten days of training in advance of the implementation of the new curriculum. In the event, they only received five days of training (MasterCard Foundation 2016). Research evidence shows that this lack of training is problematic: a study from Tanzania (Paulo 2014) showed that, even when pre-service teachers had knowledge of teaching methods appropriate to a competence-based curriculum, they did not adopt

them in their practice, as their knowledge was mainly theoretical and superficial. Similarly, in Rwanda, researchers found that, although teacher trainers view learning as knowledge construction and broadly subscribe to a learner-centred pedagogy, they often fail to use this pedagogy in their own practice due to constraints on the resources available and difficulties in sharing innovative methods with their colleagues (Peeraer et al 2015). These issues will undoubtedly lead to their trainees also failing to use this pedagogy. As the curriculum is rolled out to each year group over the next two years, then, it can be anticipated that many teachers will be delivering a curriculum for which their training has not adequately prepared them. Any Teaching Evaluation System (TES) developed for this context will need to be sensitive to this gap between what is expected of teachers and what their training equips them to deliver, and should include an element of CPD for teachers.

School-Based Mentoring Programme

There are 28,785 secondary school teaching staff in the country, of whom 19,923 (69.2%) are qualified (i.e. they hold at least a two-year teaching diploma for lower secondary or a bachelor's degree for upper secondary). However, some rural secondary schools have up to 80% unqualified teachers. This high percentage of un- or under-qualified teachers in secondary schools is a reflection of the success of the Government's efforts to increase access to education, and has created a need for intensive training for these non-qualified but serving teachers (Rutaisire and Gahima 2009). To this end, a School-Based Mentoring Programme (SMBP) was established in 2012, specifically to improve English language proficiency (as the national language of instruction had recently changed to English) and the pedagogical skills of in-service teachers. To aid delivery of this, teachers were recruited from neighbouring Uganda. The SMBP was re-designed in 2016 in line with the demands of a new competence-based curriculum (see above). The aim is for each all-through school to have a School Based Mentor (SBM), recruited for English language and pedagogic skills, who will have a reduced timetable and provide peer learning and support, specifically for the teaching of the new curriculum and to support English language proficiency. It may be possible to build some elements of this system in to any TES designed for Rwanda.

Other challenges and opportunities

As stated above, SETI aims to improve teacher retention and motivation – both of which are highly problematic in Rwanda. Motivation and morale among secondary school teachers is fragile and low, with 40% of surveyed teachers agreeing that 'teachers at my school are increasingly de-motivated' (Bennell and Ntagaramba 2008). These poor levels of motivation can be ascribed to low pay (although pay is much higher in the secondary sector than in the primary sector) and poor working conditions. Teacher workloads are constantly increasing, partly as a result of increased enrolments but also due

to the implementation of new initiatives such as the competence-based curriculum. Pay in rural schools is often even lower than in urban schools, as it is often supplemented by parental contributions through school Parent Teacher Associations, which are lower in rural areas where poverty levels are higher. Teachers in rural areas also find it more difficult to supplement their income by engaging in other employment or business opportunities. Even in urban areas, teaching is not seen as a profession of first choice for graduates, with the result that most top graduates choose to follow other career paths. In recognition of poor motivation and in an attempt to attract more graduates into the profession, the government has recently introduced incentives including supplying teachers with cows, laptops and tuition fees for their children, and the setting up of a low interest credit scheme (Republic of Rwanda 2013, World Bank 2011), although these initiatives are yet to be evaluated. Any TES introduced into this context, then, needs to be extremely sensitive to issues relating to motivation. In order to be equitable, it should also take into consideration the impact on schools in different contexts – specifically, to ensure that rural schools and teachers are not further disadvantaged by the system in any way.

Lessons from elsewhere in sub-Saharan Africa

Given that Teaching Evaluation Systems are particularly sensitive to the context in which they are used, it is instructive to examine the experience of implementing Teaching Evaluation Systems in countries that bear some similarities to Rwanda. Unfortunately, research literature into such contexts is hard to come by. Much of the data relating to student achievement in sub-Saharan Africa (SSA) is cross-sectional rather than longitudinal, which makes it more difficult to identify teacher or other factors that contribute to improved outcomes, and findings of educational effectiveness research from elsewhere is arguably much less relevant as it does not take into account the much more difficult working conditions there (for example, much larger class sizes) (Azigwe et al 2016). Limited research into mathematics attainment in Ghana, though (ibid.), suggests that the impact of teachers on student outcomes may be even higher in low income than high income countries. Additionally, there are a small number of valuable examples of various types of TESs which can be learnt from. These are detailed below.

The rapid and highly successful expansion of basic education in **Kenya**, like Rwanda (see above), has resulted in a teaching workforce where many teachers are under- or unqualified (Wanzare 2002). Government reviews in the late 1980s and late 1990s and a research study in 2002 (ibid.) into teacher evaluation found that logistical problems such as a lack of funds and poor transport availability, along with ‘incompetent, untrained and inexperienced personnel’ (ibid. p217), meant that government inspections of schools and teachers were not effective. In response, a system of School-Based Teacher

Evaluation (SBTE), including an element of Peer Teacher Evaluation (PTE), was introduced. PTE has many positive features as an element of a TES (see TES evaluation instruments section, below), including the way in which it facilitates professional dialogue between teachers and produces benefits for both the evaluator and the evaluatee (Arnodah 2013), although Arnodah makes it clear that awareness of the purpose and training in the conduct of PTE is critical to its successful implementation. The research conducted (ibid.) in thirty-seven secondary schools of different types in three districts of Western Province showed that PTE systems lacked these critical features. In the minority of schools that used PTE, only 8% of teachers had received INSET in its conduct, and even those who had received training were unclear about its' purpose. Most only used the analysis of the professional tools of teachers (schemes of work, lesson guides, records of work covered, learners' exercise books) as evidence, rather than gathering evidence from a range of instruments, and many were under the impression that this analysis was done to fulfil an administrative accountability responsibility rather than to help teachers improve their practice. Arnodah and Wanzare both put this failure down to a lack of a clear teacher standards and evaluation policy, which led to schools not having the knowledge or skills to engage in any meaningful practice. This example illustrates both the need for clear direction from policy makers and the need for proper training of all participants in the purpose and conduct of any evaluation system, in order for it to be effective.

In **South Africa**, lack of agreement between different stakeholders, particularly the teachers' trade union, has led to a confused picture (CDE 2015). The current system, the Integrated Quality Management System (IQMS), was developed to ensure quality public education by integrating the individual developmental appraisal of teachers (to ascertain CPD needs), performance measurement (to make judgements relating to pay progression) and whole school evaluation. It was implemented in the mid-1990s (ibid., Education Labour Relations Council 2003), and included formative and summative objectives. Evidence is gathered using a lesson observation schedule and teacher interviews. However, it now lacks credibility with both teachers and principals, as the professional development element that was supposed to be built in to the system has been neglected, and the system was open to accusations of bias due to one of the two assessors being chosen by the appraised teacher. A new system, the Quality Management System (QMS) was developed by the Education Labour Relations Council (ELRC) to provide a mechanism for evaluating performance levels and a basis for salary progression, but did not include any professional development component. As of 2015, its use had still not been approved by the teachers' trade union; as a result, the discredited IQMS was still being used. A 2016 qualitative study of five schools in South Africa (Whitley 2016) confirmed these findings and also revealed that schools are now designing and implementing their own appraisal systems to use alongside the IQMS system, in order to address its shortcomings. Whitley found that

‘the overwhelming perception of the IQMS is that it is a time-consuming process that does not contribute anything positive to the development and experiences of teachers’ (ibid. p90). School principals, who have to administer the system, do not support it as they were not involved in the design phase. In contrast, the school-designed systems were more successful as they were ‘tailor made to fit the specific school’ and ‘gained the support of teachers’ (ibid. p89). Indeed, these systems are so successful that the results are used by school leaders in strategic planning. The issues encountered here show the importance of meaningful consultation with all stakeholders when designing a Teaching Evaluation System, and the need for transparency and lack of bias in order for any system to have credibility.

Until very recently, **Malawi** had no coherent teacher standards or evaluation system (Chimombo et al 2014). However, in 2015, with inputs from Scottish and German NGOs, a new set of National Education Standards was introduced, which was clearly underpinned by Teacher and Educational Effectiveness Research and which now provides the framework for school and teacher evaluations (MoEST 2015a). Inspections now look at outcomes for students, the teaching process and the leadership and management of a school, with improvement plans being written as the outcome. These improvement plans include the identification of CPD needs of teachers (MoEST 2015b). Although the inspection system is designed for whole school appraisal, rather than for the formal appraisal of individual teachers (indeed, the document in which the standards are set out states that they ‘are not designed to be used to judge the work of an individual teacher’ (MoEST 2015a p10)), it includes eight standards which specifically relate to the teaching process, including curriculum knowledge, classroom management, inclusive practices and constructive use of assessment (see table 1, below).

The teaching process	7	A curriculum which is appropriate and relevant
	8	High expectations
	9	Teachers with good professional, subject and curriculum knowledge
	10	Well-planned lessons
	11	Teaching for effective learning
	12	Accurate and constructive use of assessment
	13	Teaching which meets the needs of all students
	14	Effective management of behaviour

Table 1. The Education Standards – Teaching Process. Source: MoEST 2015a

For each standard, there are four Levels of Achievement ranging from 'Below Minimum Standards' to 'Effective Practice'. This ensures that it can be applied to teachers at any stage of their career. There is also a parallel advisory system, in which advisors use the standards to support schools as they strive to improve, and may evaluate individual teachers against the standards as part of this process (ibid.). This very new system currently lacks evaluation, but the design appears to be coherent and meet the requirements of a potentially effective system, such as the formative nature of the scheme, the use of a set of clearly defined standards which all teachers are expected to meet, the clear link to professional development needs, and the use of a range of sources of evidence. How effective it is remains to be seen.

In **Zambia**, teacher appraisal and CPD needs are separated (Doyle et al 2017). Teachers are appraised as part of the appraisal process for all public service workers; there is no specific form for teachers, and, therefore, no compulsory use made of evidence gathering instruments such as lesson observations or student outcomes (although presumably these could be used as part of the appraisal process). There is room for 'further action' to be suggested, where any CPD needs could be addressed, but this is not a specific outcome of the process. CPD in Zambia is generally delivered through the Lesson Study model, introduced by JICA and adapted for the Zambian context, with mixed results. Where investment (financial and time) has been concentrated, mainly in STEM subjects, teachers are positive about the system. In other cases, though, especially where teachers are expected to use their own time for Lesson Study activities, it has been less successful and has led to some resentment. This case illustrates the need for any system to be implemented equitably across all affected groups.

The situation in **Tanzania** is similar to that in Zambia, in that teachers are appraised as public servants, with no link to CPD or the school inspection process (Doyle et al 2017, Matete 2016). School inspectors observe lessons and analyse documents, and are expected to advise teachers on their practice in any feedback, but this is not part of the teacher appraisal process. Schools should be inspected annually, but budget constraints meant that only 23% of schools were visited in 2013. CPD is, again, focussed mainly on teachers of STEM subjects, with teachers of other subjects receiving none or very little. The appraisal system was introduced in the early 2000s as part of a Public Service Reform Programme which was implemented to improve public service delivery, and takes the form of a contract between every public servant and his/her immediate supervisor (i.e. between teachers and school heads, in the case of schools). The employee has to be assured of the availability of the resources necessary to facilitate their performance, presumably to ensure the setting of realistic, achievable targets (a key characteristic of any sound appraisal process). However, one study (Matete 2016) showed that almost half of the teachers surveyed had refused to sign their contract, for a range of reasons. Some felt that their working conditions (e.g. large class sizes, lack of resources) meant that the government was not

fulfilling their side of the contract, while others felt that headteachers implemented it in a biased way. Still others felt that the system was not suitable for the context of education. These issues give more weight to the argument that teacher appraisal systems need to be contextualised and agreed between all stakeholders.

Research from **Botswana**, although outdated, illustrates the problems that can occur when there is a disconnect between government policy on teacher appraisal and what actually happens in practice (Monyatsi, Steyn and Kamper 2006). The appraisal system was introduced in 1992 and was designed to be open and transparent with a clear emphasis on formative assessment. However, by the early 2000s, when this research was conducted, many teachers were either distrustful of the scheme (a majority failing to disagree that the system was abused by Senior Management Teams) or stated that it was only there for appearances' sake and made no material difference to their practice. This research shows the importance of vigilance in implementing any system and the importance of clearly informing each new cohort of teachers of the aims and objectives of teacher appraisal.

Most recently, the World Bank, in partnership with the African Economic Research Consortium and the African Development Bank, has developed Service Development Indicators which are a set of metrics used for benchmarking service delivery performance in health care and primary education. They assemble objective and quantitative information so that users of these services can hold the providers accountable. The instruments used to collect data relating to teachers are very similar to the Stallings classroom observation instrument (see below), and are available at http://siteresources.worldbank.org/AFRICAEXT/Resources/SDI_instruments_Kenya.pdf. The data gives information about the characteristics of schools rather than the quality of teaching, and have not been used in the secondary education context, so have limited relevance for the SETI project.

Literature review

The literature review begins with an exploration of the concept of teacher quality, and the important role that it is now known to play in student outcomes (both academic and social). It goes on to summarise previous literature reviews relating to Teaching Evaluation Systems (TES), particularly with regards to the different purposes of different systems. A number of existing sets of standards and frameworks at a range of scales and across different locations are reviewed, and their applicability or otherwise to the SSA context is discussed.

Why teacher quality?

Since the 1980s, research has focussed on identifying teacher behaviours that are positively related to student achievement and value-added progress. This was as a result of research findings that

showed that ‘within school variation...dwarfs between school variation’ (Reynolds et al 2003), including UK research that ‘a pupil taught by the most effective teacher would achieve an end-of-year test score 20 per cent higher than a pupil taught by the least effective teacher’, (ibid. 2003 p92) and that ‘the relationship between teacher behaviors and pupil achievement gains was strongest in those schools with the most deprived intake’ (ibid. 2003 p92). Researchers began to realise that their focus should shift from just individual schools to both schools and individual teachers within those schools, and, in particular, what those teachers were doing in their classrooms. Teddlie et al (2003) asserted that teacher evaluation is an essential part of an effective education system and needs to be closely linked to School Effectiveness Research, Teacher Effectiveness Research, staff development, teacher improvement and school improvement, with the best systems having standards which are based on Teacher Effectiveness Research, evaluation processes which assess the process of teaching according to these standards and student outcomes, and staff development which arises out of the evaluation process and leads to teacher and therefore school improvement.

However, the vast majority of research has been conducted in the context of high income countries, including the identification of those behaviours which are associated with better student outcomes. Exporting these attributes to the SSA context, where there are typically large class sizes and scarce resources, is highly problematic, and there is a strong need to focus on adapting different approaches to teacher evaluation to the priorities and needs of specific country contexts (Crossley 2010).

Westbrook et al (2013) in their rigorous literature review of pedagogical practice in low income countries, identified six interactive and communicative practices commonly used by teachers which led to positive learning outcomes for their students. These were:

- ‘demonstration, explanation drawing on sound pedagogical content knowledge;
- flexible use of whole-class, group and pair work where students discuss a shared task;
- frequent and relevant use of learning materials beyond the textbook;
- open and closed questioning, expanding responses, encouraging student questioning;
- use of local languages and code switching;
- planning and varying lesson sequences’ (ibid. p37).

A TES which looks for these teaching practices may be more successful than one which sets standards that are so ambitious as to mark most teachers as failing.

International literature

The international literature regarding TESs has addressed the issue of teacher performance evaluation at different times, educational levels and contexts. In the following section, five of the most recent reviews of international practices in this field are summarised. It was decided to include studies no older than 10 years to reflect on the cutting-edge evidence relating to the topic. There are older reviews such as: (Clark, 1993; Colby, Bradshaw, & Joyner, 2002; L. Darling-Hammond, Wise, & Pease, 1983). However, the most recent developments present updated knowledge. Reviews were selected if they included information relating to:

- Local and/or national case studies where TESs have been implemented.
- Guidelines for effective TESs implementation.
- A range of teacher evaluation instruments that could be adapted in various contexts.

Examination of the literature will provide evidence of best practices regarding TES implementation to inform the MCF about how to pursue their goals in Rwanda's context.

Previous literature reviews relating to Teaching Evaluation Systems

Isoré, M. (2009) *Teacher Evaluation: Current Practices in OECD Countries and a Literature Review. OECD Education Working Papers, No 23.* Paris, France: OECD Publishing

This paper analyses current practices regarding TESs in primary and secondary education in different OECD countries, including various local case studies from the US context, Finland, England, Chile and France. This research includes a chapter on the purposes of evaluation, for example summative and formative goals. The advantages and disadvantages of teacher evaluation implementation are also addressed e.g., the costs and the effects of these policies on the curriculum and teaching culture. The pros and cons of the use of students' academic outcomes as a proxy for teacher quality are also regarded. In her review, Isoré points out the contentious scenario in which teacher evaluation as a means for teacher and student improvement is embedded, stressing the inconclusiveness of current academic evidence to support this idea. The report stresses how difficult it is for policy makers to decide which instruments to include in their evaluation system. In that sense, a range of teacher evaluation methods that are used for teacher assessment are presented: classroom observation, interviews of the teacher, portfolio prepared by the teacher, student outcomes, teacher test, and questionnaires to different stakeholders who interact with the teacher on a regular basis, e.g. parents and students. Isoré stresses that there are multiple stakeholders in the establishment of a Teaching

Evaluation System, however she asserts that ‘an effective, fair and reliable evaluation scheme requires teachers’ overall acceptance and appropriation of the system’ (ibid, p31). This is a critical feature of effective TESs that will be revisited in the recommendations of this report.

Darling-Hammond, L. (2012). *Creating a Comprehensive System for Evaluating and Supporting Effective Teaching*. Stanford, California.

Darling-Hammond is a longstanding advocate of standards-based teacher evaluation. In this review, the author proposes guidelines for the enactment of evaluation systems that support effective teaching on a local level. In her paper, the author emphasises that,

Teaching Evaluation Systems need to be considered not only in terms of evaluation instruments or procedures, but also in terms of the policy systems in which they operate and the school-based conditions that are needed to stimulate continuous learning and improvement (Darling-Hammond, 2012, p39).

Five operational elements of documented best practice regarding TESs are proposed: 1) start with standards; 2) create performance-based assessment; 3) build a standards-based system of local evaluation; 4) create structures to support high-quality, fair, and effective evaluation; and, 5) create aligned professional learning opportunities. The local case studies included in the review are:

- New Mexico’s Standards Based Teaching Evaluation System
- Standards-Based Evaluation in San Mateo, California
- Use of Student Learning Evidence in Amphitheatres, Arizona
- Massachusetts’ Multiple Measures System of Evaluation
- Peer Assistance and Review (PAR) Toledo, Ohio and the California style.

In line with other educational effectiveness research (Marzano & Toth, 2013; OECD, 2014; Santiago & Benavides, 2009) the author suggests that TESs should include multiple sources of information relating to student learning, teacher practice, and the contribution of the teacher to the wider professional community. Furthermore, she suggests that evaluators should be well trained, and that the outcome of assessment should lead to professional development. Darling-Hammond has been critical of the use of value-added measures (VAM) of student learning as the only criteria for teachers’ assessment. She argues that:

1. Value-Added models of teacher effectiveness are highly unstable
2. Teachers’ value-added ratings are significantly affected by differences in the students assigned to them

- Value-added ratings cannot disentangle the many influences on student progress (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012, pp. 9–10).

See more discussion about VAM in the Instruments section below.

OECD. (2013). *Teachers for the 21st Century: Using Evaluation to Improve Teaching*. Retrieved from <http://link.springer.com/10.1007/978-3-642-36970-4>

This report summarises existing teacher evaluation approaches in 28 countries, noting that 22 of them have policy frameworks in place for at least one type of assessment while six have none. Figure 1 presents the list of countries surveyed by the OECD including the different stages of a teacher career that may be subject of assessment.

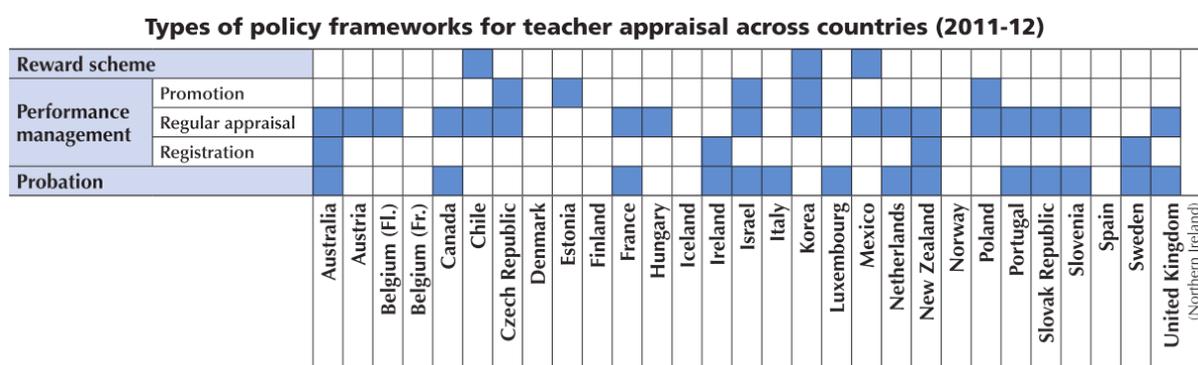


Figure 1. Source: OECD (2013).

The report also addresses issues of governance and distribution of responsibilities for appropriate implementation of TESSs. Regarding the goals of teacher evaluation, the authors suggest that combining summative and formative purposes under a single policy should be well planned because teachers are likely to be more open to show their teaching weaknesses if the purpose of assessment is for teacher development rather than accountability. On the contrary, if there are high-stakes consequences such as rewards and sanctions attached to the evaluation, teachers may be less open to disclose their areas for improvement. The OECD review suggests that when summative purposes are attached to TESSs, involving internal as well as external evaluators is advised ‘to ensure fairness across schools’ (*ibid.*, p19). The report also suggests that teacher appraisal information can be used for different purposes such as ‘hiring and tenure decisions, promotion opportunities, or, in some circumstances, responses to underperformance’ (*ibid.*). It is worth noting in this report that countries like Finland and Sweden lack working standards or frameworks for teacher evaluation. In these

contexts, assessment is a responsibility of school leaders and is commonly conducted in annual reviews or conversations between the school principal and the teacher.

The most common evaluation instruments identified in this report are: classroom observation, interview/dialogue between a school leader and the teacher, teacher self-appraisal, teacher portfolio, teacher testing, and students results. Other instruments are the review of lesson plans, logs of conversations with parents, and samples of students' work.

Schmelkes, S. (2015). Assessment of teacher performance-state of affairs. In *Critical issues for formulating new teacher policies in Latin America and the Caribbean: the current debate*. UNESCO 2015.

Schmelkes' paper is a review of the literature about models of assessment which highlights the relevance of teaching standards as the cornerstone for the design and implementation of TESs. The report discusses the influence of Danielson's Framework for Teaching in the US context, as well as in some developing countries in Latin America. The report includes an updated review of current developments regarding TESs in Mexico, Cuba, Chile, Colombia and Peru.

Schmelkes discusses situations where teachers resist policies on TESs and raises concerns relating to the necessity of involving the different stakeholders (e.g. teachers) in policy formulation. It is asserted that successful TESs are those devised and supported by the teachers, such as the cases of Australia, England and Chile (*ibid.*). This paper acknowledges quality of teaching as 'the most important factor (in terms of education provision) behind differences in pupils' learning outcomes' (*ibid.*, p138). However, school leadership and contextual conditions where teachers work are also regarded as influential in the overall teacher performance. Therefore, it recommends that these factors are also considered when assessing teachers.

This report indicates the following instruments for the assessment of teachers: classroom observations, portfolios, self-assessment, evidence of student learning, teacher interviews, school principal report, knowledge tests, and peer/students/parents/ survey.

The Centre for Development and Enterprise, South Africa (2015) *Teacher Evaluation – Lessons from other countries*. Johannesburg, South Africa: Centre for Development and Enterprise

This paper was written as a response to the crisis in teaching in South Africa, in terms of both student outcomes and in the quality of teachers. It notes that an effective TES could help to improve the quality of teaching in South African schools, and seeks to identify the most successful practices in TESs across a range of countries. Evidence was collected from reports published by the OECD, the Bill and Melinda Gates Foundation, the New Teacher Project and the National Commission of Teaching and America's

future, i.e. mainly high income highly industrialised contexts. Having reviewed the evidence, the paper states that the common trends in TESs in these countries are:

- The use of a set of professional standards, formulated with the involvement of the teaching profession, as a benchmark.
- The use of multiple sources of evidence for making judgements about teachers, with lesson observations being used most commonly, student surveys and peer-assessment, and measures of student outcomes – although all of these need to be designed and used with great care to ensure reliability and validity.
- High quality professional development opportunities which are closely aligned with the TESs.
- A significant investment, in terms of human capital, finance, and time, to ensure that the system is the correct one for the context.

While recognising the potential of a well-designed and implemented TESs to make a real contribution to improving teacher effectiveness, this paper also acknowledges the risk of a poorly designed and implemented system encouraging teachers of low-performing students or in low-performing schools to move schools or leave teaching altogether – which would have the devastating effect of increasing educational inequality.

Summary of findings from previous literature reviews

The international literature on TESs presented previously has shed light on some of the contexts where teacher evaluation policies have been implemented. As noted, TESs can be found in a range of settings, from local to national policies that the whole teacher workforce must observe. Importantly, TESs can be found in developed, as well as in developing countries. The international experience points out the best practices to implement TESs by providing important benchmarks that a developing system such as the Rwanda's can learn from. It has been noted that effective TESs:

- **Involve teachers and other stakeholders during formulation:** doing so can help to reduce resistance during implementation and gain support for it.
- **Are clear on their purposes and scope:** it is important to define whether the system seeks the development of teachers, if it will be used for accountability purposes or both.
- **Are guided by Standards or Frameworks of good teaching:** most countries that have a Teaching Evaluation System have their own, or an adaptation of standards and/or frameworks which orients the design of methods and instruments for teacher assessment.
- **Use multiple sources to collect evidence regarding teachers' quality/effectiveness:** there are a range of instruments and methods that are recurrent in the literature. These instruments will be reviewed in more detail below.

- **Provide teachers with useful information on their areas for improvement:** and most importantly, these weaknesses are tackled via Continuing Professional Development (CPD).
- **Are well planned from the outset:** this means that ways of integrating summative and formative purposes have been considered. Also, that the governance of the evaluation system has been made clear to all parties.
- **Acknowledge the prevalence of context in the students' outcomes:** not just teachers, but other factors around the classroom have an influence on teaching and learning.
- **Take account of the positive and the negative consequences:** since teacher evaluations can affect the whole education system in many ways, it is important to ponder the advantages and disadvantages of implementation.

Below the purposes of TESs, a definition of Teacher Quality, and current developments regarding Standards and Frameworks for Teaching will be reviewed.

Literature relating to Teaching Evaluation System goals and purposes

It is widely agreed that when designing or upgrading a Teaching Evaluation System the definition of goals or purposes of the evaluation system should be decided (Goe, Holdheide, & Miller, 2014; Santiago & Benavides, 2009). In this regard, the two most common purposes documented in the literature are for accountability and for teacher development, or what is called summative and formative evaluation respectively (Donaldson & Papay, 2014). Some TESs pursue one of the two, such as the cases of Finland (Sahlberg, 2006, 2011) and the province of Fukoka, Japan (Martinez, Taut, & Schaaf, 2016), where teachers' practice reviews are only used to inform Continuing Professional Development (CPD) paths. Most models of TES pursue these two intentions all-in-one, i.e. the appraisal of teachers' performance for accountability purposes, including rewards and sanctions, but with a commitment to improving the teachers' practices via CPD. It should be noted, though, that there are some dissenting voices regarding this practice: Popham (1988) argued that the tensions between formative and summative aims render any system that has both to be unworkable, and Murphy (2013) asserts that the summative and formative aspects of any system should be separated as far as possible to ensure that discussions about strengths and weaknesses can be frank and honest. Harmonizing these two elements of a TES is the most challenging part that policy makers and implementers have encountered.

An essential conceptual differentiation

It is important to make a conceptual differentiation between teacher quality and teaching quality. Darling-Hammond presents the following:

***Teacher quality** might be thought of as the bundle of personal traits, skills, and understandings an individual brings to teaching, including dispositions to behave in certain ways. **Teaching quality** refers to strong instruction that enables a wide range of students to learn (Darling-Hammond, 2012, pi).*

Goe (2007) elaborates more on this matter. In a review of the literature Goe proposes to look at teacher quality considering three aspects: *inputs*, *processes* and *outcomes* as follows:

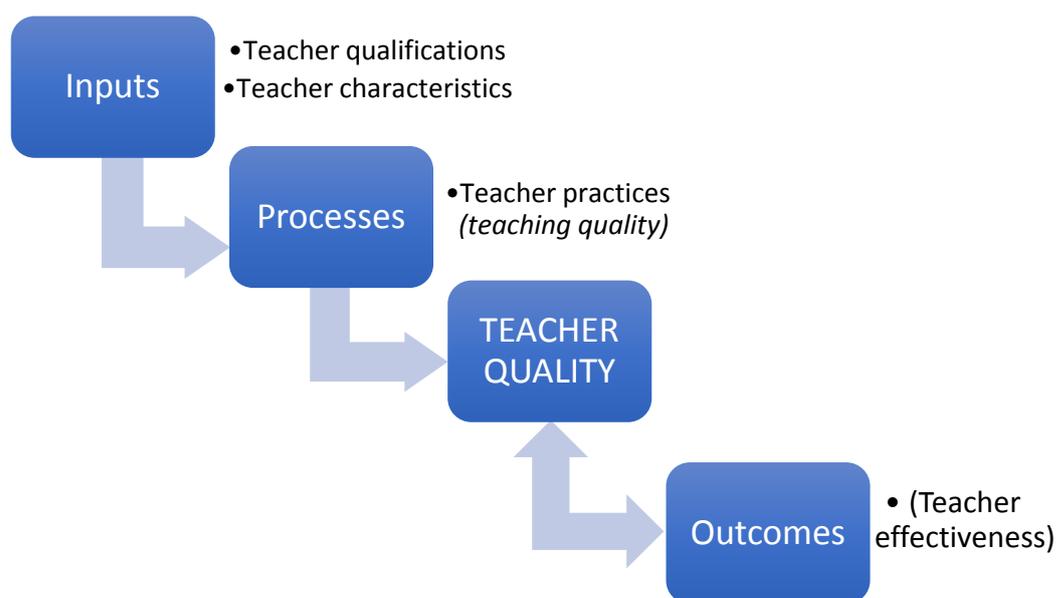


Figure 2. *Teacher quality and components. Adapted from (Goe, 2007).*

Having established this differentiation, when teacher assessment refers to qualifications (e.g. work experience), characteristics (e.g. beliefs), or classroom performance (e.g. pedagogy), the qualities which are being judged are the teachers' *inputs* and/or *processes* of their practice (Goe, 2007). Under this definition teacher quality exists independently of students' outcomes. On the other hand, *teacher effectiveness* is more frequently found in the literature that looks at the teachers' impact on students'

outcomes in its various forms and measures. The review conducted by Goe was focused only on the literature addressing students' gain in growth models or Valued-Added Models of teacher effectiveness (*ibid.*). However, teacher effectiveness remains a contentious construct because teachers certainly impact on a broader set of students' outcomes not necessarily represented in tests scores (Goe, Bell and Little, 2008).

For the purposes of this report, the term 'teacher quality' will be used to refer to both the characteristics of individual teachers and the processes and practice they employ in their teaching. This report will not explore teacher effectiveness, as measured by student outcomes, because the mechanisms for tracking student outcomes and linking them to individual teachers are not currently available in Rwanda or in other countries in which MCF is working. Nevertheless, the potential use and relevance of student learning and value added progress measures, as an important element of any TES, will be returned to in the recommendations.

'Standards' and 'Frameworks' for Teaching

In this report, 'standards' refer to statements describing what is expected of a teacher's knowledge and performance in their day-to-day teaching, developed for guidance or in order to make a judgement about those teachers. They may be minimum standards, which need to be achieved to gain entry to the teaching profession, or they may be for use with more experienced teachers to, for example, identify and reward excellence. 'Frameworks', for the purposes of this report, are the mechanisms used in order to reach judgements about whether a teacher has attained the standards or not. A framework includes an element of progression for teachers as they improve their practice. Frameworks may include sets of standards for teachers at different stages of their careers (for example, for gaining qualified teacher status initially and then for being accredited as an advanced skills teacher), but they also set out the pathway for moving from one stage to the next.

Literature relating to Standards for Teaching

Once the goals for a TES have been decided, the establishment of standards is the next step to take. Coherence between student learning standards and standards for teaching is crucial (Linda Darling-Hammond, 2012; Draper, Hofmeyr, & Hohnston, 2017). Furthermore, defining who a 'good teacher' or an 'effective teacher' is, or what could be regarded as 'good teaching' can serve as the basis for formulation of the different elements enclosed in the standards (Linda Darling-Hammond, 2012; Goe et al., 2014; Schmelkes, 2015). The most common models of standards are:

- *Professional profiles of teachers or teaching standards (general profile of competencies for teachers), including specialised profiles for particular types of teachers (e.g. level of education, subject);*
- *A set of general and professional duties of teachers, including job descriptions; and*
- *At the school level, a school development plan, internal regulations and the annual activity plan. (OECD, 2013).*

In the following section, standards and frameworks from a range of international contexts will be presented. These examples were selected based on their salience in the academic literature, and the existence of evidence of their applicability to contexts similar to the ones that MCF works in (i.e. lower income countries). Some of them have been widely used and adapted in local, as well as in national settings.

Standards used in the US

In the US the National Board for Professional Teaching Standards (NBPTS, 2017) and the updated Interstate Teacher Assessment and Support Consortium (InTASC) (Council of Chief State School Officers, 2011) are two prominent references in terms of expected teachers' knowledge and performance. The former framework is more generally used for certification of accomplished teaching, whilst the latter has been used to evaluate classroom performance of teachers at various career stages. Both sets of standards present criteria of what can be regarded as best teaching practice. The NBPTS¹ is organised in propositions at the more general level (5 items), followed by standards (10 items), and the teachers dispositions towards students which describe in detail the expected teacher's practice (NBPTS, 2001). The INTASC standards (10 items) are accompanied by indicators which 'are examples of how a teacher might [...] demonstrate each standard' (Council of Chief State School Officers, 2011, p6). The indicators are grouped in three categories: performances, essential knowledge, and critical dispositions. Performances stand for 'the aspect that can be observed and assessed in teaching practice' (*ibid.*, p6); essential knowledge is the 'declarative and procedural knowledge as necessary for effective practice' (*ibid.*); finally, critical dispositions involve habits and moral commitments to the professional activity of teaching. Table 2 summarises the NBPTS and the INTASC standards.

¹ This example regards the Early Adolescence through Young Adulthood/ Art Standards, for teachers of students ages 11–18+

NBPTS	InTASC
Five core propositions	Four categories (10 standards)
Teachers are committed to students and their learning	The learner and learning (3 standards)
Teachers know the subjects they teach and how to teach those subjects to students	Content knowledge (2 standards)
Teachers are responsible for managing and monitoring student learning	Instructional practice (3 standards)
Teachers think systematically about their practice and learn from experience	Professional responsibility (2 standards)
Teachers are members of learning communities	

Table 2. Source: (Council of Chief State School Officers, 2011; NBPTS, 2017).

Australian Standards

The Australian Standards are meant to ‘articulate what teachers are expected to know and be able to do at four career stages: Graduate, Proficient, Highly Accomplished and Lead’ (Education Services Australia, 2011, p1). The descriptors of the standards convey the insight of nearly 6,000 teachers. The standards are expected to boost better teacher practices and students’ outcomes. Also, they are intended to present a common ground for conversations between teachers, and contribute to the enhancement of their profession. The standards were envisaged to be used ‘as the basis for a professional accountability model’ (*ibid.*, p2).

Table 3 presents The Australian Standards.

Domains of teaching	Standards
Professional Knowledge	<ol style="list-style-type: none"> 1. Know students and how they learn 2. Know the content and how to teach it
Professional Practice	<ol style="list-style-type: none"> 3. Plan for and implement effective teaching and learning 4. Create and maintain supportive and safe learning environments 5. Assess, provide feedback and report on student learning
Professional Engagement	<ol style="list-style-type: none"> 6. Engage in professional learning 7. Engage professionally with colleagues, parents/carers and the community

Table 3. Source: (*ibid.*, p3).

Each of the seven standards include a variable number of *focus areas* (37 items in total) that provide a clearer description of what is expected of a teacher's work. The Australian standards stand out because these are designed for four teacher career stages (i.e. Graduate, Proficient, Highly Accomplished and Lead), rather than issuing a one-size-fits-all policy.

Teacher standards in England

Until 2012, there were three sets of standards for teachers in England, according to the different stages of a teacher's career: a set relating to the awarding of Qualified Teacher Status, core standards for teachers on the main pay scale, and post-threshold standards for teachers who had been teaching for six years or more. Together, these standards formed a framework which reflected the progression that was expected of teachers as they moved through their careers (Goepel 2012).

However, in 2012 these sets of standards were replaced by a single set, to be applied to teachers at any stage of their career (*ibid.*). There are nine standards, eight of which relate to teaching and one relating to professional conduct, as follows:

Part one: Teaching

1. Set high expectations which inspire, motivate and challenge pupils.
2. Promote good progress and outcomes by pupils.
3. Demonstrate good subject and curriculum knowledge.
4. Plan and teach well structured lessons.
5. Adapt teaching to respond to the strengths and needs of all pupils.
6. Make accurate and productive use of assessment.
7. Manage behaviour effectively to ensure a good and safe learning environment.
8. Fulfil wider professional responsibilities.

Part two: Personal and professional conduct.

A teacher is expected to demonstrate consistently high standards of personal and professional conduct.

Each standard is sub-divided into bullet points which serve to amplify the scope of the standard in question. This new system appears to have removed the element of progression. However, in reality, schools have adapted these new standards to develop their own frameworks which set out both threshold performance levels and the levels expected as teachers gain more experience (see appendix B for an example).

An overview of existing Frameworks for Teaching

Based on the standards, frameworks for teaching can be designed. Frameworks provide methods to attest a teacher's performance by describing different levels of achievement for each component of the framework. There are many examples of Frameworks for teaching documented in different contexts. Each framework here included will be critically reviewed to provide recommendations on its adaptability in Rwanda considering aspects such as scope and feasibility in that context.

The Framework for Teaching (Danielson, 2007)

Danielson's Framework for Teaching (FfT) is one of the most widely documented guidelines for teacher performance evaluation available (Goe, 2007; Papay, 2012; Santelices, Valencia, Gonzalez, & Taut, 2016; Taylor & Tyler, 2012). The framework strives to address the 'teacher's responsibilities that have been documented through empirical studies and theoretical research as promoting improved student learning' (*ibid.*, p1). This framework is well aligned with the ten standards of the former version of the INTASC standards. The FfT is organised in domains (4 items), components (22 items) and elements (76 items); table 4 presents the FfT.

Domain	Components
1. Planning and Preparation	<ul style="list-style-type: none"> a) Demonstrating Knowledge of Content and Pedagogy b) Demonstrating Knowledge of Students c) Setting Instructional Outcomes d) Demonstrating Knowledge of Resources e) Designing Coherent Instruction f) Designing Student Assessment
2. The Classroom Environment	<ul style="list-style-type: none"> a) Creating an Environment of Respect and Rapport b) Establishing a Culture for Learning c) Managing Classroom Procedures d) Managing Student Behaviour e) Organising Physical Space
3. Instruction	<ul style="list-style-type: none"> a) Communicating with Students b) Using Questioning and Discussion Techniques c) Engaging Students in Learning d) Using Assessment in Instruction e) Demonstrating Flexibility and Responsiveness
4. Professional Responsibilities	<ul style="list-style-type: none"> a) Reflecting on Teaching b) Maintaining Accurate Records c) Communicating with Families d) Participating in a Professional Community e) Growing and Developing Professionally f) Showing Professionalism

Table 4. The Framework for Teaching. Source: adapted from (Danielson, 2007).

Danielson recommends the FFT as a *road map* for preparation of new teachers, hiring and regular assessment. The FFT may also foster professional conversations amongst educators. One particularity of this framework is that it is generic, that is, it can be used for generalist teachers, as well as for subject-specific educators from different contexts and situations. Danielson argues that the FFT addresses the 'commonalities' of a teacher's work (*ibid.*, p22) such as promoting school environments where 'students feel respected by the teacher and their peers' (*ibid.*).

The FFT includes a series of rubrics to observe the teacher achievement of the different elements contained in the framework. There are four levels of performance: unsatisfactory, basic, proficient, and distinguished. The main means for assessment put forward are: classroom observations accompanied by pre-conference and post-conference meetings between the evaluator and the teacher, and data collection via artefacts such as teachers' plans and students' learning evidence.

More information about the Danielson FFT can be found at: <http://www.danielsongroup.org>.

Strengths and weaknesses: The FFT in its original form is essentially a developmental method for teacher performance assessment. Therefore, its use in Rwanda can provide implementers and teachers with relevant information on the teachers' quality which is one of the aims of MCF. However, the FFT demands that evaluators and teachers are knowledgeable of the rubrics and means of data collection such as classroom visits and artefacts scrutiny. Raters must be well-trained which can be costly and time-consuming, especially if there is not a well-established inspectorate structure, or if school principals are not ready to undertake this responsibility immediately. Use of video-taped lessons that are assessed in a different place and at a later date can be a way of addressing the issue of lack of training at school-level; however this method also has time and monetary implications that should be considered in advance. Finally, reviewing video-taped lessons and artefacts far away from where they happen can negatively affect any judgements made, given the absence of direct contextual information.

The Good Teaching Framework (Chile); The Good Teacher Performance Framework (Peru); and The Teaching Profile (Mexico).

As documented in the literature, the FFT has been very influential not only within the US context, but in other places such as Chile (Santiago & Benavides, 2009; Taut, Santelices, Araya, & Manzi, 2010; Tornero & Taut, 2010), Peru (Guzmán, 2017; Vázquez Cruz, Cordero Arroyo, & Leyva Barajas, 2014),

and Mexico (Cortez, unpublished). These three Latin American countries share similarities in their frameworks for teacher evaluation as it is presented in table 5.

	Chile	Peru	Mexico
Name of the framework	The Good Teaching Framework	The Good Teacher Performance Framework	The Teaching Profile²
Level 1	Domains 1. Preparation for Teaching 2. Creating an Appropriate Environment for Learning 3. Teaching for all Students Learning 4. Professional Responsibilities	Domains 1. Preparation for the students' learning 2. Teaching for the students' learning 3. Participation in school management linking with the community 4. Professionalization and teacher identity	Dimensions A teacher that... 1. Knows her students, how they learn, and what they should learn 2. Organises and evaluates the educational work and makes pertinent didactic interventions 3. Recognises herself as a professional who continuously improves to help the students to learn 4. Takes the legal and ethical responsibilities of her profession 5. Participates in the efficient functioning of the school linking the community
Level 2	Criteria (20 items) Example: <i>Knows the characteristics, knowledge and experiences of her students.</i>	Competencies (9 items) Example: <i>Knows and understands the characteristics of her students and contexts; knows the content, the teaching approaches and processes.</i>	Parameters (16 items) Example: <i>Knows the processes of development and learning of the teenagers.</i>
Level 3	Descriptors (70 items) Example: <i>Knows different learning styles.</i>	Performances (40 items) Example: <i>Demonstrates knowledge and understanding of the students' sociocultural</i>	Indicators (68 items) Example: <i>Acknowledges that family, society and culture are factors that have an impact on the teenagers' learning.</i>

² For secondary school teachers in the school year 2017-2018.

		<i>characteristics, and their specific needs.</i>	
--	--	---------------------------------------------------	--

Table 5. Three Latin American frameworks for teacher evaluation. Sources: Chile (MINEDUC, 2008); Peru (Ministry of Education, n.d., 2013, 2017a, 2017b; Vázquez Cruz et al., 2014); Mexico (DOF, 2013; INEE, 2016; SEP, 2017).

Strengths and weaknesses: The cases listed above illustrate some developing countries where a seminal Frameworks for Teaching (Danielson, 2007) has been adapted, so may be more instructive when considering frameworks for the context in which MCF is working. Having a road map rather than none is much more appropriate for a system that seeks to assess teachers. However, it is important to note that Standards, as well as Frameworks ought to consider the specific needs and potential constraints for implementation in each context. Foremost, all stakeholders should be able to have their say regarding what is expected from their functions and the ways it could best be attested.

Goals and Roles Performance Evaluation Model© (short title: Goals and Roles Model©)
(Stronge, 2010)

The Goals and Roles Model© developed by Dr James Stronge is aimed at improving student learning and teachers practice by collecting evidence of a teacher's performance. The model, which is used by The American Association of Schools in South America (AASSA), consists of six performance standards (general categories) and a flexible number of performance indicators which may serve as examples of observable, tangible behaviours. The six performance standards are:

1. Instructional Planning
2. Instructional Delivery
3. Assessment of/for learning
4. Learning Environment
5. Professionalism
6. Student Progress

Data on a teacher's performance is collected via various mediums and are thought to provide meaningful feedback to the teacher about their performance.

- a) Goal Setting for Student Progress
- b) Observations
- c) Teacher Documentation Folder

d) Student Surveys

The definition of goals for student progress are advised to follow the SMART³ acronym to review the worth as well as the feasibility of the ambitions stated for group of students in a school year.

Under the Goals and Roles Model© performance standards and not indicators are evaluated. That is, the more general categories of the framework are the ones that are given rates of achievement. Using the rubric as a guidance for assessment, four possible outcomes from evaluation are prescribed: unacceptable, developing/needs improvement, proficient, and exemplary. This framework is suggested for use within schools and is linked to professional development via conversations with evaluators, and the design of CPD plans.

Strengths and weaknesses: This framework emphasises the importance of goal setting and planning. It incorporates guidelines to design feasible goals and notably, involves the students' insight regarding teaching delivery. The framework has been prescribed to be implemented in lower income Latin American countries. However, it is specific to American schools in the different contexts, which might not necessarily reflect the general characteristics of the country of implementation.

The Australian Teacher Performance and Development Framework

(Education Services Australia, 2012)

Following recent developments, teacher evaluation in Australia has evolved into a more flexible and school situated approach to teacher performance assessment. The authorities in this context seek to foster a culture of performance and development focusing on a series of elements inherent to the teaching profession. Figure 3 illustrates the Australian framework.

³ S (specific) M (measurable) A (appropriate) R (realistic) T (time limited).

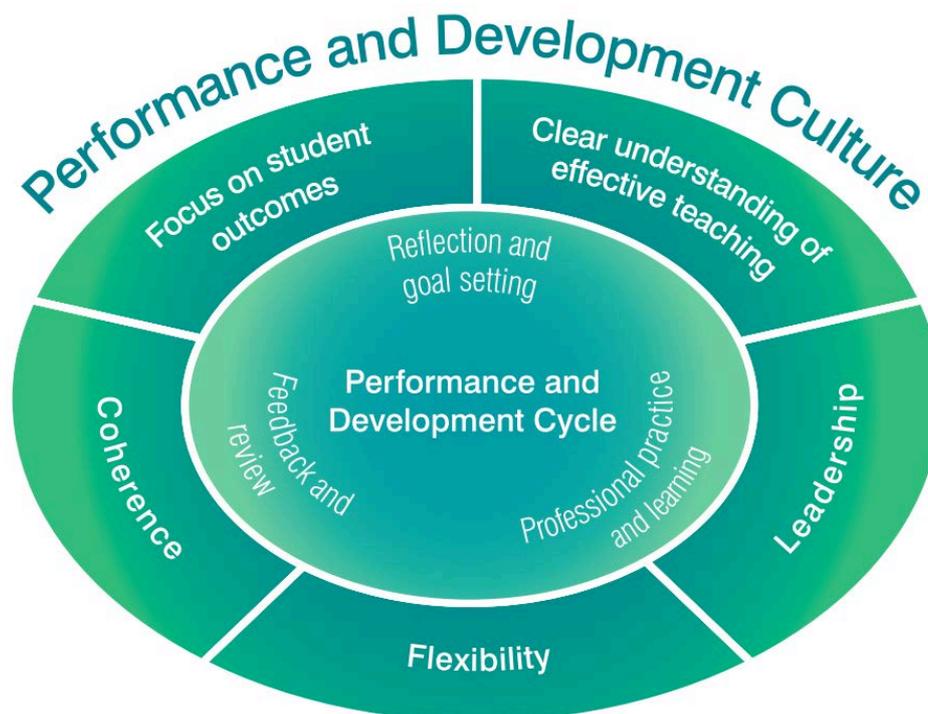


Figure 3: *The Australian Teacher Performance and Development Framework*. Source: (Education Services Australia, 2012, p5)

In Australia, there is a commitment not only to improve teaching but foremost to impact on students' outcomes. To do so, the Australian Professional Standards for Teachers are referred to as the criteria of 'what teachers should know and be able to do' (*ibid.*, p3). These standards are available from https://www.aitsl.edu.au/docs/default-source/apst-resources/australian_professional_standard_for_teachers_final.pdf. Under this framework, educators at all levels are expected to exert leadership skills for their own, and their community's enhancement. The framework requests consistency between the school's agenda and the teachers' performance and development activities. As said, this is a more flexible approach to teacher performance assessment that considers the different starting points in terms of development that characterises each school.

School principals or delegates are expected to document regular appraisals and develop different forms to measure teacher performance and development. The minimum data expected for teacher evaluation are: evidence of 'impact on student outcomes; information based on direct observation of teaching; and evidence of collaboration with colleagues' (*ibid.*, p6). Verbal as well as written formal and informal feedback for teachers at least annually is also advised.

Strengths and weaknesses: The Australian framework can be described as flexible and developmental. It focuses on the advance of a culture of good teaching with an emphasis on student outcomes. Notably, the Australian framework devolves responsibility to the schools, including the teacher, to monitor assessment of the teachers' practices. This framework which considers the different stages in a teacher's career is coherent with what Rwanda in its law is seeking to achieve (MINEDUC, 2007, p. 15). Still, the model posed in Australia seems more appropriate in a context where certain conditions for successful implementation are already in place. For instance, school principals and other school leaders enjoy acceptance from their colleagues to make judgements on their professional practices, teachers are certified to do their job, and protocols to collect and assess students' outcomes are already in place. These conditions may not exist in Rwanda or other contexts in which MCF is working, and this may make this implementation of a similar system more challenging.

The dynamic model of educational effectiveness

This model (Creemers and Kyriakides 2010) has been developed using evidence from Educational Effectiveness Research that student achievement is affected by factors operating at four different levels: student, classroom (including teacher), school and system. Emphasis is placed on the role of the teaching and learning that go on, but account is also taken of school- and system- level factors and the role they play in creating a learning environment. The model includes eight teaching/classroom factors based on the most significant teaching skills, similar to standards, and suggests five different levels of skill that teachers can progress through, similar to a framework (Antonioniou, Kyriakides and Creemers 2015).

Strengths and weaknesses: The Dynamic Model (DA) is comprehensive, that is, it provides guidelines for all levels in an education system, which goes beyond the scope and purposes of the MCF-SETI in Rwanda. However, the elements contained in the five levels of teachers' skills may inform the specific standards and frameworks based on which MCF can design the most suitable TES for their purposes.

Summary of findings from existing standards and frameworks

Frameworks and standards to evaluate teachers are certainly important for various reasons. They:

- 1) serve as a guidance of what is considered good teaching;
- 2) are the foundation of the evaluation instruments used to assess teachers, and
- 3) allow for comparison between teachers' performances at one point in time, or over a particular period (Hobson et al, 2010; Isoré, 2009; OECD, 2013).

The Standards and Frameworks listed in the previous sections are not exhaustive, but rather illustrative of the variety of options currently used in various countries.

Subject-specific frameworks

It is worth observing that most Standards and Frameworks currently available are meant to apply across generalist teachers, as well as subject-specific educators. However, a limited number of subject-specific frameworks for teacher performance assessment have been developed for the US context. For instance, the Quality of Science Teaching (QST), the UTeach Teacher Observation Protocol (UTOP), the Mathematical Quality of Instruction (MQI), and the Protocol for Language Art Teaching Observation (PLATO) are examples of observational teacher assessment systems that have been explored empirically (Kane & Douglas, 2012; Martínez Rizo, 2016). Details about each of these frameworks are given below. These tailored frameworks seem to be more aligned with SETI and might be useful for the purposes of MCF in Rwanda's secondary education institutions. However, it would be appropriate to recommend that MCF consider the use of these kind of frameworks only when the minimum conditions are in place; for example, once all evaluatees have undertaken CPD in the subject they teach, and when there is sufficient capacity in terms of evaluators and other logistical aspects that facilitate the introduction of classroom observations that are focused on specific traits of subject-specific quality teachers. These frameworks, then, could be considered for further stages of the MCF-SETI programme in Rwanda and other SSA contexts.

The Quality Science Teaching (QST)

The Quality Science Teaching (QST) is a classroom observation instrument integrated by six domains:

1. Assessing Teacher's Content Knowledge and Pedagogy
2. Engaging Students in Learning Science
3. Facilitating Scientific Discourse and Reasoning
4. Promoting Laboratory-Based Inquiry
5. Providing Opportunities for the Application of Science
6. Monitoring Student Learning

(Schultz & Pecheone, 2014, pp447–448).

Eighteen indicators and several quality indicators are assessed using a four-point Likert scale (1 to 4). The rubrics pertaining to the QST can be consulted in Schultz and Pecheone, (2014, p484). Appendix C illustrates domain 1 and the QST-MET cluster 1. The creators of the QST assert that:

The selection of the domains and indicators was informed by and aligned with the National Science Education Standards (NRC, 1996), Inquiry and the National Science Education Standards (NRC, 2000), What Teachers Should Know and Be Able to Do (NBPTS, 2002), and Taking Science to School: Learning and Teaching Science in Grades K–8 (NRC, 2007) (*ibid*, p448).

Since some of the members of the QST team participated in the formulation of the Next Generation Science Standards (NGSS)⁴, there is an overlap between the two to a certain degree.

An adapted version of the QST was used in the Measures of Effective Teaching project (the QST-MET) to evaluate biology teachers' quality practices. Some of the non-official findings from this project which involved a total of fifty video-taped lessons were that:

- Teachers formulate low-order questions to the students, that is, factual, rather than more profound reflections are common among teachers in their sample. (Domain 3).
- There were no lessons where the teachers promote the applicability of knowledge to the reality. (Domain 5).

The findings above are referred as non-official because 'the QST-MET official version consisted of three domains or clusters, with a total of twelve indicators' (ibid, p450). Also, the modified instrument included a zero "0" rate when a specific action was not observed. The modification was made mainly thinking about the raters' cognitive demand that the original QST would require from them. The three QST-MET clusters are:

1. Assessing Teacher's Knowledge and Pedagogy
2. Engaging Students in Science Learning
3. Promoting Laboratory-Based Inquiry

Teachers who volunteered to submit video lessons had to provide four videos per school year with at least one lab lesson. Findings from the project reveal a low, rather significant positive correlation between lessons in the lab and outside (0.304, $p < 0.01$), with lab lessons gaining lower scores. This indicates that teachers' practices differ in one setting and the other, which can inform training for quality instruction. Overall, the QST-MET reveals that biology teachers in their sample have low content and pedagogical knowledge, as well as scientific practices, and urges actions to improve this situation.

The UTeach Teacher Observation Protocol (UTOP)

The UTOP is a classroom observation instrument focused on mathematics and sciences teachers practices (Anonymous, n.d.). The instrument was originally designed to attest the preparedness of student teachers who undertake the UTeach, a programme that grants a degree in mathematics or science, and a certification for teaching. According to documentation:

⁴ Consult the Next Generation Science Standards here: <https://www.nextgenscience.org>

The UTOP (see UTOP, 2009) includes 32 classroom observation indicators organized into four sections: Classroom Environment, Lesson Structure, Implementation, and Math/Science Content. The indicators are rated by observers on a 7-point scale: 1 to 5 Likert with Don't Know (DK) and Not Applicable (NA) options (for some items). Each of the four sections concludes with a synthesis rating on a 1 to 5 Likert scale, which is intended to be a summary or overall score based upon the observer's weighting of the relative importance of the indicator evidence. The UTOP also includes a post- observation teacher interview, which contains 12 open-ended questions to assist the observers in ascertaining the background and purposes of the lesson. Finally, the full UTOP includes both observer and teacher demographic forms, as well as spaces for the observer to fill in background information about the lesson (Anonymous, n.d., p12).

The protocol can be used across different education levels, from preschool to the undergraduate level (The UTeach Institute, n.d.). The protocol was used in the MET project which involved '982 videos of grades 4 to 8 mathematics classrooms' (Marder & Walkington, 2014, p234). One of the conclusions of the study was that value-added measures and information coming from observations that used the UTOP provide complementary and separate insight 'on what happens in classrooms' (ibid, p235). The MET project revealed that teachers who scored high in the UTOP would also get high value-added scores; however, the UTOP is inconclusive in distinguishing teachers with low and middle value-added rates. Marder and Walkington (2014) present a response to the issue of imprecision that is worth reviewing. See observation protocol in appendix D.

The Mathematical Quality of Instruction (MQI)

The Mathematical Quality of Instruction (MQI) was designed by Hill, Rowan, and Ball (2005, cited in Hill et al., 2008) in order to quantify the relationship between Mathematical Knowledge for Teaching (MKT) and teacher quality of instruction (Hill et al., 2008). Hill and colleagues have explored in different studies their observational teacher evaluation instrument and students' outcomes, claiming the MQI as a valid and reliable tool to detect effective teachers. The elements that the MQI seeks to appraise in a mathematics lesson are presented in table 6.

Elements of Mathematical Quality of Instruction

-
- Mathematics errors—the presence of computational, linguistic, representational, or other mathematical errors in instruction;
 - Contains subcategory specifically for errors with mathematical language
 - Responding to students inappropriately—the degree to which teacher either misinterprets or, in the case of student misunderstanding, fails to respond to student utterance;
 - Connecting classroom practice to mathematics—the degree to which classroom practice is connected to important and worthwhile mathematical ideas and procedures as opposed to either non-mathematical focus, such as classroom management, or activities that do not require mathematical thinking, such as students following directions to cut, color, and paste, but with no obvious connections between these activities and mathematical meaning(s);
 - Richness of the mathematics—the use of multiple representations, linking among representations, mathematical explanation and justification, and explicitness around mathematical practices such as proof and reasoning;
 - Responding to students appropriately—the degree to which teacher can correctly interpret students' mathematical utterances and address student misunderstandings;
 - Mathematical language—the density of accurate mathematical language in instruction, the use of language to clearly convey mathematical ideas, as well as any explicit discussion of the use of mathematical language.
-

Table 6. Source: (Hill et al., 2008)

Being an observational tool, the MQI utilises the rubric as a screening instrument and rates teachers in a 3-point scale in each of the above-mentioned elements (Kane & Douglas, 2012).

Instruments currently used for Teaching Evaluation Systems

Systems of teacher evaluation around the world use a wide range of instruments to reach their judgements (see above). It is very important that the instruments chosen are appropriate to the country, priorities and context in which they are being used, and that they measure what they are intended to measure, i.e. they are carefully matched to the purpose of the assessment. Designers of teaching quality assessment schemes must ensure that they are not driven by 'wrong' drivers – for example, they should not choose instruments purely for ease of use, or because they are a good fit with the technology currently being used – but that they are making considered decisions based on research into the strengths and weaknesses of the instruments. The following is a review of findings relating to different possible instruments from research by Goe, Bell, and Little (2008), CDE (2015), Isoré (2009), Murphy (2013), and Papay (2012), some of which are themselves syntheses of earlier research. Consideration will also be made of the suitability of each instrument in the context of SSA.

Classroom observations

These are the most common sources of evidence used in OECD countries (CDE, 2015, Isoré, 2009). They are generally supported by teachers, and are extremely useful for ascertaining their CPD needs, which may make them particularly useful in the SSA context (CDE, 2015), although they do not appear to correlate particularly closely with gains in student test scores (Murphy, 2013). They allow for constructive feedback to teachers, which other sources of evidence do not make provision for, and this has been shown to improve effectiveness of teachers in the long run (Murphy, 2013). Teddlie and Reynolds (2000) assert that quality of instruction can only be assessed through direct observation. Classroom observations can be conducted in person or videoed. Two examples of the latter case are the video component of the teacher portfolio in the NBPTS certification of accomplished teaching, and the lessons prepared and videotaped for teacher evaluation in Chile. Video-taped observations have also been used for the International Mathematics and Science Study (TIMSS), and the Measures of Effective Teaching (MET) project (Martínez Rizo, 2016). Video recorded lessons may be particularly useful instruments in low income countries with poor accessibility, especially to rural areas.

A review of the literature (Martinez, Taut, & Schaaf, 2016), analysed 16 cases in six countries where classroom observations are used as part of teacher performance management. It is noted that teachers' content knowledge, their ability to plan lessons, take account of their students' characteristics and how to evaluate learning are recurrent elements that are assessed using observations across TES in their sample. Still, Martínez et al. (2016) stress that teacher classroom observation is an approach to teacher performance management that applies differently in different contexts and situations. They distinguish three main differing elements: conceptual aspects, methodological issues, and context. Among the conceptual aspects that make classroom observations different between one setting and another are: Standards or frameworks for teaching, the focus of the observation (for example, instruction, planning, etc.), target population (for example, novice, experienced teachers, etc.), weight of evaluation within the overall TES, and the role of observation in further CPD. With regard to the methodological issues, differences can be found in the frequency of observation, who the observers are and crucially whether they are peers (who may, for example, focus on development) or internal/external experts/inspectors (who are more likely to focus on accountability), whether announced or unannounced observations are planned, the training provided to evaluators, and research on the quality of the observations (e.g. reliability and validity). Finally, the context dimension is concerned with the accountability model in place, the locus of control (e.g., national or school level, etc.), and the stakes for participant teachers. The resource-heavy nature of classroom observation as an instrument, in terms of training of observers (see below) and time for observations and feedback, may make it more problematic in the SSA context.

Reliability issues

Reliability is a major concern with classroom observations (Papay, 2012), and, due to these concerns, some researchers assert that they should not be given too much weighting in any formal assessment processes (Murphy, 2013). In order to increase reliability, it is recommended that classroom observations are carried out by more than one (Kane & Douglas, 2012) well-trained internal and external evaluators (to reduce the possibility of biased results) and that each evaluation incorporates multiple observations carried out over a period of time (Darling-Hammond, 2012). In contrast, Marzano and Toth (2013) suggest that four observations throughout a school year are still limited as far as drawing conclusions on a teacher's performance is concerned, and points out the conservative reliability scores reported in the MET project which are about .30. Inter-rater reliability (i.e. the difference between two or more evaluators), sampling error (i.e. the difference between different observations), and measurement error (i.e. inaccuracy in terms of what is evaluated and the level of complexity used in a teaching task) are latent in any observation protocol.

Validity issues

By comparing teacher effectiveness ratings by principals and teacher value-added measures, Harris, Ingle and Rutledge (2014) found that validity is also contentious as school leaders tend to focus on the teachers' characteristics and their contribution to the school community as well as instruction practices when assessing them using observations. For example, teachers who isolate from the rest of the teachers tend to receive lower rates by their principals as compared with their value-added measures. Although principals can provide relevant information on a teacher's practice from an insider's perspective, it is important to consider that school principals may reach their judgements based on a broader picture of the teachers' performance, and not only in terms of their impact on students' outcomes. Validity issues were apparent in the South African, Tanzanian and Botswana cases cited above (CDE 2015, Matete 2016, Monyatsi, Steyn and Kamper 2006), where evaluatees became distrustful of the system at least in part because observers were internal and perceived to be biased.

Martínez et al., (2016) noted an important difference regarding what is being observed and appraised between their Asian and American case studies. While Japan and Singapore frameworks seek to observe teachers' attitudes and competencies towards 'nurturing the whole child, winning hearts and minds; pupils' values and human rights, [and] pupil trust' (p21-22), the American cases focus on the teachers' instructional abilities such as 'questioning techniques, classroom management, [and] identify progress for subgroups' (p22). Therefore, classroom observations should be coherent with the evaluation system and the culture of implementation. Consultation with all stakeholders to achieve agreement on what is being observed and appraised is critical.

Examples of systems of lesson observation

Two examples of systems of lesson observation have been evaluated in terms of their suitability for Low Income Countries (LICs) (Bruns, De Gregorio and Taut, 2016). These were the Classroom Assessment Scoring System (CLASS) and the Stallings classroom observation instrument.

In the CLASS, teachers are scored on eleven dimensions of their practice, categorised into three domains: emotional support, classroom organisation and instructional support. Observers use rubrics to look for very specific behaviours in each dimension. The system gives a multi-dimensional measure of teacher quality and has been found to be valid in a number of studies. However, it is very expensive to use (US\$90-100 per teacher) and it requires use of well-trained observers. This makes it problematic for use in LICs.

The Stallings classroom observation instrument is relatively simple, with teachers being observed at ten equally spaced points throughout a lesson and coded according to the activity they and the class are engaged in. There is an Open Data Kit software version of the instrument which can be used on tablets (available at <http://www.worldbank.org/en/programs/sief-trust-fund/brief/stallings-program-download>), and it generates quantitative data related to the teacher's use of classtime which can then be compared across and between settings. Observers require very little training and results between observers show consistency.

The study found that the Stallings instrument was probably more appropriate for LIC settings, both because of the low cost and because it measures baseline skills which must be present for instructional quality to improve. However, they found that it is 'too crude to be used for individual teacher performance evaluations' (Bruns, De Gregorio and Taut, 2016, p28), but better suited to large scale studies aiming to evaluate the impact of new education programs.

Two other observation instruments were evaluated as part of the Effective Classroom Practice Project (Sammons and Ko, 2008) which focussed on exploring typical and more effective classroom practice of teachers in English primary and secondary schools. These were the International System for Teacher Observation and Feedback (ISTOF), produced by teams from 21 participating country experts, and the Lesson Observation Form for Evaluating the Quality of Teaching, based on an international collaboration between English and Dutch inspectorates. Examples of the observation schedules can be found in the appendices of the report. The evaluation found that differences in context (such as the level of social disadvantage) did not exert a strong effect on teachers' practice, providing support for the view that 'effective teachers share many common features' (ibid p83). They also found that

some of these features are shared across different sectors, subjects and contexts. However, as this study was conducted in only English schools, it may have limited value for the context of LICs.

Advantages: classroom observations are welcomed by teachers, especially if for teacher development purposes rather than accountability (see literature review); can be performed by school authorities, as well as external evaluators; can provide useful feedback on the teacher's classroom practice for improvement purposes.

Disadvantages: Classroom observations are conducted by human beings; therefore, reliability and validity can be an issue. In order to minimise reliability and validity issues, several observations by more than one well-trained evaluator are recommended, which adds to the high cost of implementation, both in terms of time and finance.

Teacher interview/dialogue/appraisal with colleagues and school authorities

One of the oldest approaches to teacher appraisal is the formal and informal dialogue between the teacher and the school principal. In the US, the Praxis III system has used this method since 1988 (Martínez Rizo, 2016). Occasional conferences with teachers is a characteristic of education systems where inspectorates are well established, such as in England (Santiago & Benavides, 2009), and are commonly coupled with classroom observations (Darling-Hammond, Wise, & Pease, 1983). In Chile, teachers are co-evaluated by a peer colleague using a structured interview which accounts for 20% of the overall evaluation (Docentemás, 2017). Using multilevel modelling, Taut et al. (2014) found a positive, rather weak correlation between teachers results in their interview with students' value-added estimates in Spanish and Mathematics. As noted, teachers' interviews are currently used in various contexts and can be conducted by school leaders and colleagues alike. Given its closeness with observations, this method for teacher assessment is better suited to provide meaningful and readily available feedback to the teacher. This feedback is valuable in contexts such as Rwanda, where the formative element of teacher evaluation is foregrounded. However, evaluators need to be well-trained if problems relating to using this tool, as seen in the example from Kenya, are to be avoided. Such training may be expensive.

Advantages: Can be performed by internal, as well as external evaluators; can be useful in providing teachers with feedback on their teaching practice. Can be incorporated into existing school management structures and systems.

Disadvantages: It requires time allocation for conferencing; some structure is needed to provide the teacher with meaningful feedback, which in some cases will involve some training and preparation (see above).

Teacher testing

This is very rarely used. In most contexts where it is used, the purpose is to ascertain the skills of newly qualified or prospective teachers. For example, in the UK professional skills tests are used to check that candidates for teacher training are sufficiently qualified in numeracy and literacy to enter the profession, regardless of their specialism (CDE, 2015). However, teacher testing is also used for other purposes, for example, teachers seeking a certification from the NBPTS have to sit a content-knowledge assessment in a secure centre (NBPTS, 2017). In Chile, teachers who participate in the rewards programme called AVDI present a content and pedagogical knowledge exam (MINEDUC, 2017). Following the most recent Mexican Education Reform, teacher testing has been implemented at various teacher career stages in Mexico; for instance, there are examinations to enter to the service; to be granted tenure at the end of the first and second year in the profession; as mandatory regular appraisal every four years; to pursue posts as school leader, Technical and Pedagogical Advisor, Schools' Supervisor; and for a national teacher reward scheme (DOF, 2013).

In the Rwandan context where there are large numbers of un- and under-qualified teachers, testing may be appropriate as a diagnostic tool, to establish whether teachers are in need of CPD to increase their content knowledge. Once tests are developed, they can be relatively cheap and straightforward to implement.

Advantages: can be informative of the content-knowledge of teachers; when open-ended items are part of it, teacher tests can be useful to investigate the teachers' critical reflection on a given topic. Relatively resource-light.

Disadvantages: can be difficult to address the teacher's pedagogical knowledge although some tests have attempted this end. Given its standardised nature, tests cannot address the contextual differences in which teachers operate. Teachers may not be willing to participate.

Peer assessment

This is another instrument that is particularly effective in ascertaining CPD needs and delivering CPD to teachers (CDE, 2015; Muijs & Reynolds, 2005), although it is not particularly useful as an instrument for assessing the quality of individual teachers. It is often used as part of a programme where teams of teachers plan lessons and sequences of lessons collaboratively, and then observe each other's delivery of these lessons, as in the Japanese Lesson Study programme (Muijs & Reynolds 2005). The School-Based Mentoring Programme in Rwanda already includes an element of peer assessment, so this could be usefully built on. The Zambian, South African and Kenyan examples, though, show the

importance of high quality training for peer assessors and the provision of time in which assessment can be conducted and feedback given.

Advantages: it is essentially formative therefore, teachers can be more open to show their professional needs in order to get meaningful feedback on their practice. Can add to the school culture of improvement.

Disadvantages: cannot be immediately implemented in the absence of peer-observation culture. It requires well-established leadership and collegiality. Time allocation for classroom observation has to be negotiated to avoid disruption in the provision of education service to the students of teachers who take the role of observers.

Teacher portfolios

These are collections of materials assembled by teachers as evidence of their teaching practices, school activities and student progress (Goe, Bell and Little 2008). They can provide comprehensive portraits of individual teachers and allow for a broader definition of effective teaching, as they are not restricted to evidence from the classroom or student outcomes. Portfolios, however, can be adapted to the needs of the general TES, for instance, in the NBPTS, currently an e-portfolio is required as part of the evaluation. Broadly speaking, teachers must include three entries (video and students' work samples) (NBPTS, 2017). In Chile, the portfolio includes 1) Description of an 8 hours plan and the response of a questionnaire, and 2) Video of a 40-minute lesson (Docentemás, 2017). Until 2016 in Mexico, teachers had to elaborate a teacher portfolio that included 1) Four samples of students' work; two samples regarding literacy, and two samples of mathematics from the top and bottom students in terms of proficiency level, and 2) A reflexion on the students' work including the context, the didactic strategies used, and the assessment and feedback given to the students (SEP, 2016). Teachers generally welcome the use of portfolios as part of performance management (Villafuerte, Quinto Simón, Navarro Corona, Leyva Barajas, & González Nava, 2015). However, they can be very time-consuming to compile and there are problems related to reliability and validity in their assessment. Where teacher motivation and morale are low, as in Rwanda and much of SSA, teachers would be unlikely to dedicate time and effort to putting together such a portfolio unless they were convinced of the benefits they would receive as a result.

Advantages: they give the teachers agency in deciding what products to show externally as a result of their work. Portfolios are welcomed as a means for teacher performance management because the kind of elements included relate to the teachers' professional activities.

Disadvantages: since portfolios have to be assembled for evaluation purposes, teachers have to devote time to elaborate them. Reliability and validity can be an issue when it comes to assessment of portfolios. Time is needed to evaluate the portfolio.

Classroom artefacts

This is a relatively recently developed instrument, with infrequent use. It refers to the examination of artefacts such as mark books, lesson plans and student work, in order to assess teacher quality. It has the advantage of being less demanding on time and resources than full lesson observations while still providing evidence of the work being undertaken in the classroom. The artefacts have already been created by the teacher or students, rather than being put together specifically, as with a portfolio. Danielson, 2007, suggests using her Framework for Teaching together with classroom observations and review of teaching-learning artefacts. In the NBPTS teachers complete written reflections on evidence of their students' work. Similarly, in the recently enacted Mexican Teacher Evaluation, teachers are expected to write down some reflections about samples of their students' work and upload them onto a dedicated online hub. However, there are no existing structured protocols to compare the artefacts from different teachers, so reliability and validity cannot be assessed. Without clear direction from policy makers and line managers, evaluation of artefacts may be reduced to a box-ticking exercise, as in the Kenyan case study.

Advantages: time-allocation is not required as artefacts naturally occur in the everyday teaching; can be reviewed by the teachers as well as other evaluators, such as school leaders or external evaluators.

Disadvantages: there is no standardised way to evaluate artefacts, which compromises reliability and validity.

Student surveys

Students have the most experience of individual teachers and could be expected, therefore, to be able to return valuable evaluations of their practice (Goe, Bell and Little 2008). Indeed, research has shown that student surveys are reliable and provide useful insights for teachers (CDE, 2015, Isoré, 2009). However, researchers caution that, as students are not trained to rate teachers and may not value the qualities which are supposed to enhance students' learning, the results of such surveys should only be used alongside other evidence and not as the sole or primary evaluation criterion (Goe, Bell and Little 2008, Isoré 2009). As student surveys are relatively resource-light, they may form a useful element of a TES in Rwanda and similar contexts.

Advantages: students can provide useful information about how teachers interact with them and how they organise their learning activities. Due to limitations maybe best used for formative purposes.

Disadvantages: students can express their views on a teacher's practice that do not necessarily refer to teaching quality or teacher effectiveness. Surveys require thorough piloting and need to be adapted to the students' level of reading and understanding to use them as a valid source of information. Time and other resources are needed to complete the evaluation and interpretation.

Teacher self-assessment

This instrument enables teachers to reflect on their own performance, identifying strengths and weaknesses, and is, therefore, valuable when considering CPD needs, which is a priority in the case of Rwanda and other SSA countries (CDE, 2015). It can provide insights that other instruments cannot, as the subjects are the only ones with full knowledge of their own abilities and needs (Goe, Bell and Little 2008). However, it should not be used in isolation, especially in the case of high stakes assessments. Research into the validity and reliability of such assessments has produced mixed results.

Advantages: depending on the stakes, teachers can be very open in stating their professional needs; self-evaluation is most times welcomed by teachers because it regards their own strengths and weaknesses in the profession.

Disadvantages: where the stakes are high, teachers may not express their teaching weaknesses. Due to so-called social desirability, even if the evaluation is for formative purposes, teachers might not express all their needs.

Value-added student outcomes

These are teacher effectiveness measures, usually based on standardised academic subject tests, that are calculated by taking into account a variety of factors that may contribute to student performance, most crucially prior attainment but also including other relevant factors outside the control of the school such as parental income and level of parental education. At their simplest, they measure students relative progress in a class or school (in comparison to students in other schools) from a known baseline (rather than simply raw attainment scores). Their use in TESs assumes that the students of high quality teachers will make the most progress.

Supporters of value-added student outcomes argue that they focus directly on student learning and are relatively objective (CDE, 2015, Isoré, 2009). In addition, recent developments in technology have ensured that the statistical models used to measure student progress and link it to specific teachers are reliable (Goe, Bell and Little 2008). Teacher Effectiveness Research (previously mentioned above) shows that the results of individual teachers taken over a number of years can reveal the underlying impact of these teachers, while other research indicates that these measures are valuable in identifying the best teachers and those who are consistently performing badly, but not so good at

differentiating between the majority of teachers who perform just above or below average (Murphy, 2013).

In spite of considerable use of teacher effectiveness measures in some contexts such as the USA, in other contexts the data is often not available, student samples are not large enough within individual teacher classes or the data is not reliable enough for such measures to be used, especially as the sole or main evidence base for judging teacher quality. This is especially the case for subjects outside the core of literacy and numeracy, and in lower-income countries. In addition, student outcomes do not take account of other ways in which teachers contribute to the development of their students (CDE, 2015). These measures also fail to determine CPD needs, as they do not reveal which teaching methods or practices contribute the most to learning gains (Goe, Bell and Little 2008). Although there is much to recommend these measures, then, they are currently inappropriate for the SSA context due to data issues. If these issues were to be resolved, they may be a useful tool as part of a TES in the future.

Advantages: are considered a fairer way to appraise the contribution of teachers to the students' learning as compared with raw tests results; can help to identify the top and bottom performers in a whole district or education system.

Disadvantages: current value-added teacher effectiveness models are not perfect, and should not be used as the sole source of information for decisions on individual teachers; implementation is costly and requires well-established protocols and the use of unique identifiers to link data records from students, teachers and schools; currently a limited number of subjects can be studied using value-added models; there are many more aspects that have been critiqued about VAM for teachers that can be consulted (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Mortimore et al., 1994; Papay, 2012; Sørensen, 2016).

Concluding remarks relating to instrument selection and design

This review demonstrates clearly that there is no one-size-fits-all instrument for assessing teacher quality. The important considerations when choosing instruments are the resources available to the assessors (and the constraints on these resources), the context in which the system will be used, and the purpose of the system. Systems with both formative and summative purposes will need to make use of a range of instruments.

Recommendations

The following recommendations arise from the evidence and findings of all the previous sections, taking into consideration the key principles guiding the SETI and the context in which the TES will be implemented.

Over-arching principles for the establishment of a TES:

- The governance of the TES should be explicit from the start. Most successful TES are designed in a collaborative manner, with inputs from all stakeholders, including teachers and school leaders. This collaborative design will facilitate teacher buy-in to the process from the beginning.
- The TES must be customised to align with the context and culture in which it is being used. This principle means it is not appropriate to take an existing TES and implement it in a new context.
- Any policy must be evidence-based, drawing on academic research and good practice in existing TESs, especially from low income countries and SSA.
- The TES should be part of an integrated system, drawing on Teacher Effectiveness Research, which includes evaluation processes and methods which assess the process of teaching according to agreed standards and student outcomes, and staff development which arises out of the evaluation process, with an overall purpose of teacher and therefore school improvement.
- The TES should align with other related education policy and planned reforms, in order to avoid overlap or conflicting aims and purposes.
- In Rwanda, the TES should potentially have formative assessment as its primary focus, given the context of low morale and poor retention rates among teachers, although there is also a need for it to have a summative role for monitoring and accountability purposes. However, the risks that arise from potential tensions between formative improvement purposes of teacher evaluation and summative accountability purposes need to be explicitly recognised in the TES.

The establishment of a TES can be viewed as a process. The following recommendations are for each stage of the process. It should be noted that designing the system is not necessarily a linear process, but may be iterative, as each step is informed by and informs the steps before and after it.

Stage 1 – establishing teaching standards

- Teaching standards need to be appropriate to the context in which the TES is being used. They should take account of not only the characteristics of teachers but the constraints under which they are working, such as the large class sizes and lack of resources that are often characteristic of schools and classrooms in SSA. This will ensure that any judgement based on the standards will be meaningful to educators.

Stage 2 – developing a framework

- The framework needs to be based on the teaching standards.
- It should allow for progress in relation to the standards, so that it can be applied to teachers at any stage of their career.
- The framework should provide for identification of CPD needs, to facilitate teacher improvement in relation to the standards.
- The framework may also indicate approaches, methods and instruments of teacher evaluation.

Stage 3 – developing instruments for gathering evidence and making judgements

- The TES should make use of a range of instruments and sources of evidence including but not limited to lesson observations. Instruments which enable the identification of CPD needs should be given priority, given the low income country context.
- Student progress outcomes should only be used in the TES if the data is valid and reliable, to ensure that judgements are equitable. This is currently not the case in Rwanda, but the development of a data management system may be an opportunity for future investment.
- Teacher tests may be useful as diagnostic instruments to establish CPD needs relating to content knowledge, especially for un- or under-qualified teachers.

Stage 4 – implementation of the TES

- Proper training of the individuals who will function as evaluators is vital for effective implementation. Likewise, those who will be the subject of assessment ought to know and have training relating to the standards expected of their professional activities.
- Time-allocation for teacher preparation for assessment and to participate in CPD should be considered as an integral part of TESs.
- The implementation of any TES should not add to teacher workload. In the event of there being workload implications, these must be negotiated with teachers (e.g. through their trade unions).

- The TES should align with and, as far as possible, make use of any current systems of teacher appraisal, assessment of CPD needs or performance management, as well as any broader related policies to avoid overlap or conflicting aims and purposes.
- Any TES must be applied equitably across the whole teaching body, regardless of subject taught, experience, or role within the school.
- The TES needs to be systematically evaluated and monitored over time to ensure successful implementation. It should be revisited regularly by all stakeholders to ensure that it is fit-for-purpose and that all stakeholders are clear about its aims and objectives.
- Unintended consequences of the TES must be recognised and addressed as soon as possible.

References

- Anonymous. (n.d.). *U Teach Observation Protocol*. Retrieved from https://uteach.utexas.edu/sites/default/files/UTOP_Paper_Non_Anonymous_4_3_2011.pdf
- Antoniou, P., Kyriakides, L. and Creemers, B. (2015) The Dynamic Integrated Approach to teacher professional development: rationale and main characteristics. *Teacher Development* 19:4, 535-552
- Arnodah, I. W. A. (2013) Teacher improvement through peer teacher evaluation in Kenyan Schools, *European Journal of Training and Development*, Vol. 37 Issue 7, 635-645
- Azigwe, J., Kyriakides, L., Panayiotou, A. and Creemers, B. (2016) The impact of effective teaching characteristics in promoting student achievement in Ghana. *International Journal of Educational Development*, Vol 51, 51-61
- Bennell, P. and Ntagaramba, J. (2008) *Teacher Motivations and Incentives in Rwanda: A Situational Analysis and Recommended Policy Actions*. Available at: <http://www.rencp.org/wp-content/uploads/2010/06/Teacher-Motivation-and-Incentives-in-Rwanda-08.doc> (Accessed: 7 September 2017)
- Black, P. and Wiliam, D. (2001) Inside the Black Box – Raising Standards Through Classroom Assessment. *Phi Delta Kappan* Vol 80, 2, 139-148
- Bruns, B., De Gregorio, S. and Taut, S. (2016) *Measures of Effective Teaching in Developing Countries*. RISE Working Paper 16-009
- Centre for Development and Enterprise (CDE) (2015) *Teacher Evaluation in South African Schools*. Johannesburg, South Africa: Centre for Development and Enterprise
- CDE. (2015). *Teacher Evaluation: Lessons from other countries*. Johannesburg, South Africa.
- Chimombo, J., Meke, E., Zeitlyn, B., and Lewin, K. (2014) Increasing access to secondary school education in Malawi: does private schooling deliver on its promises? *Privatisation in Education Research Initiative Working Paper No. 61*. Available at: http://www.periglobal.org/sites/periglobal.org/files/WP-61_Increasing%20Access%20to%20Secondary%20School%20Education%20in%20Malawi.pdf (accessed: November 2016).
- Clark, D. (1993). *Teacher Evaluation: A review of the literature with implications for Educators*. California State University at Long Beach.
- Colby, S. A., Bradshaw, L. K., & Joyner, R. L. (2002). *Teacher Evaluation: A Review of the Literature*. In *American Education Research Association* (p. 18). New Orleans, LA.
- Council of Chief State School Officers. (2011). *InTASC Model Core Teaching Standards: a Resource for State Dialogue*. Washington, DC.
- Creemers, B. and Kyriakides, L. (2010) School Factors Explaining Achievement on Cognitive and Affective Outcomes: Establishing a Dynamic Model of Educational Effectiveness. *Scandinavian*

Journal of Educational Research, 54:3, 263-294

Crossley, M. (2010) Context matters in educational research and international development: Learning from the small states experience. *Prospects* 40: 421-429

Danielson, C. (2007). *Enhancing professional practice : a framework for teaching*. (2nd ed.). Alexandria, Virginia: : Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2012). *Creating a Comprehensive System for Evaluating and Supporting Effective Teaching*. Stanford, California.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating Teacher evaluation. *Phi Delta Kappan*, 93(6), 8–15.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher Evaluation in the Organizational Context : A Review of the Literature. *Review of Educational Research*, 53(3), 285–328. Retrieved from <http://www.jstor.org/stable/117036>

Docentemás. (2017). Teacher Evaluation Instruments. *Website Docentemás, Chile*.

DOF. (2013). Ley General del Servicio Profesional Docente. Retrieved March 15, 2016, from http://dof.gob.mx/nota_detalle.php?codigo=5313843&fecha=11/09/2013

DOF. (2013). DECRETO por el que se expide la Ley del Instituto Nacional para la Evaluación de la Educación. Retrieved from http://www.dof.gob.mx/nota_detalle.php?codigo=5313842&fecha=11/09/2013

Donaldson, M. L., & Papay, J. P. (2014). Teacher Evaluation for Accountability and Development. In M. E. G. Helen F. Ladd (Ed.), *Handbook of Research in Education Finance and Policy* (2nd ed., pp. 174–196). Routledge. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=zCbfBQAAQBAJ&oi=fnd&pg=PA174&dq=new+teachers+more+willing+to+evaluation&ots=ZBG0mgeEHI&sig=XCMLONKawLCKoJQX6_KiNdFuss0#v=onepage&q&f=false

Doyle, H., Barrett, A. and Reeves, A. (2017) *Improving the Quality of Teaching in Secondary Education: Scoping Studies for Zambia, Malawi, Tanzania, Cote d'Ivoire and Senegal*. Bristol, UK: University of Bristol

Draper, K., Hofmeyr, J., & Hohnston, A. (2017). *TEACHER PROFESSIONAL STANDARDS FOR SOUTH AFRICA*. (A. Bernstein, Ed.). Centre for Development and Enterprise.

Education Labour Relations Council (2003) *Collective Agreement Number 8 of 2003*. Centurion, RSA: ELRC

Education Services Australia. (2011). *Australian Professional Standards for Teachers*. Retrieved from https://www.aitsl.edu.au/docs/default-source/apst-resources/australian_professional_standard_for_teachers_final.pdf

Education Services Australia. (2012). *Australian Teacher Performance and Development Framework*. Retrieved from https://www.aitsl.edu.au/docs/default-source/default-document-library/australian_teacher_performance_and_development_framework_august_2012.pdf?sfvrsn=e

7c2ec3c_0

Ellett, C. and Teddlie, C. (2003) Teacher Evaluation, Teacher Effectiveness and School Effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education* Vol 17, 1, 101-128

Goe, L. (2007). The Link Between Teacher Quality and Student Outcomes : A Research Synthesis. *Quality*, (October), 1–72.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*. Washington DC,.

Goe, L., Holdheide, L., & Miller, T. (2014). *Practical Guide to Designing Comprehensive Teacher Evaluation Systems A Tool to Assist in the Development of Teacher Evaluation Systems*.

Goepel, J. (2012) Upholding public trust: an examination of teacher professionalism and the use of Teachers' Standards in England, *Teacher Development*, 16:4, 489-505

Guzmán, E. (2017, August 29). La Evaluación del Desempeño Docente en Perú. *Online Newspaper*. Retrieved from <http://www.gacetaucayalina.com/2017/08/29/la-evaluacion-de-desempeno-docente-en-el-peru/>

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How Teacher Evaluation Methods Matter for Accountability: A Comparative Analysis of Teacher Effectiveness Ratings by Principals and Teacher Value-Added Measures. *American Educational Research Journal*, 51(1), 73–112. <http://doi.org/10.3102/0002831213517130>

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, 26(4), 430–511. <http://doi.org/10.1080/07370000802177235>

INEE. (2016). *Replantea el INEE el modelo de evaluación del desempeño profesional docente para 2017*. Mexico, Mexico.

International Task Force on Teachers for Education for All (Teacher Task Force) (2015) *Teacher Policy Development Guide – Summary*. Paris, France: UNESCO

Isoré, M. (2009). *Teacher Evaluation: Current Practices in OECD Countries and a Literature Review* (No. 23). *Education Working Papers*.

José Passos, A. F. *A comparative analysis of teacher competence and its effect on pupil performance in upper primary schools in Mozambique and other SACMEQ countries* (unpublished doctoral thesis, University of Pretoria, 2009)

Kane, T., & Douglas, O. (2012). *Measures of Effective Teaching (MET). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA.

Konstantopoulos, S. and Borman, G. D. (2011) Family Background and School Effects on Student Achievement: A Multilevel analysis of the Coleman Data. *Teacher College Record* Vol 113, 1, 97-132

Marder, C., & Walkington, M. (2014). Classroom Observation and Value-Added Models Give Complementary Information About Quality of Mathematics Teaching. In T. Kane, K. Kerr, & R. Pianta (Eds.), (First). San Francisco, CA.

Martinez, F., Taut, S., & Schaaf, K. (2016). Classroom observation for evaluating and improving teaching: An international perspective. *Studies in Educational Evaluation*, 49, 15–29.
<http://doi.org/10.1016/j.stueduc.2016.03.002>

Martínez Rizo, F. (2016). *La evaluación de docentes de educación básica. Una revisión de la experiencia internacional*. Mexico: INEE.

Marzano, R. J., & Toth, M. D. (2013). *Teacher evaluation that makes a difference: a new model for teacher growth and student achievement*. Alexandria, Virginia: ASCD.

Matete, R. E. (2016) Implementation of Management by Objective through Open Performance Review and Appraisal System for Teachers in Tanzania. *International Journal of Education and Literacy Studies*, Vol 4 (3): 24-33

MasterCard Foundation (2016) *Mapping The Secondary Teacher Environment in Rwanda*. Unpublished draft report

MINEDUC. (2008). Marco para la buena enseñanza. Retrieved January 23, 2017, from <https://www.docentemas.cl/docs/MBE2008.pdf>

MINEDUC. (2017). AVDI reward scheme. MINEDUC website. Retrieved from <http://www.avdi.mineduc.cl/Home/avdi>

Ministry of Education, Science and Technology (MoEST) (2015a) *Directorate of Inspection and Advisory Services – National Education Standards, Primary and Secondary Education*. Lilongwe, Malawi: Ministry of Education Science and Technology

Ministry of Education, Science and Technology (MoEST) (2015b) *Directorate of Inspection and Advisory Services – Handbook for Inspectors and Advisors*. Lilongwe, Malawi: Ministry of Education Science and Technology

Ministry of Education. (n.d.). *Marco de Buen Desempeño Docente*. Lima, Peru.

Ministry of Education. Reglamento de la Ley de Reforma Magisterial (2013). Preu. Retrieved from www.minedu.gob.pe

Ministry of Education. (2017a). Evaluación Docente. Retrieved August 21, 2017, from <http://evaluaciondocente.perueduca.pe/desempenoinicial/>

Ministry of Education. (2017b). Rúbricas de observación de aula para la Evaluación del Desempeño Docente. Retrieved from [http://repositorio.minedu.gob.pe/bitstream/handle/MINEDU/5332/Rúbricas de observación de aula para la evaluación del desempeño docente manual de aplicación.pdf?sequence=1&isAllowed=y](http://repositorio.minedu.gob.pe/bitstream/handle/MINEDU/5332/Rúbricas%20de%20observación%20de%20aula%20para%20la%20evaluación%20del%20desempeño%20docente%20manual%20de%20aplicación.pdf?sequence=1&isAllowed=y)

Monyatsi, P., Steyn, T. and Kamper, G. (2006) Teacher appraisal in Botswana secondary schools: a critical analysis. *South African Journal of Education* Vol 26 (2): 215-228.

Mortimore, P., Sammons, P., Thomas, S., Mortimore, P., Sammons, P. A. M., & Thomas, S. (1994). School Effectiveness and Value Added Measures School Effectiveness and Value Added Measures. *Assessment in Education: Principles, Policy & Practice* ISSN:, 1(3), 315–332.
<http://doi.org/10.1080/0969594940010307>

Muijs, D. and Reynolds, D. (2005) *Effective Teaching – Evidence and Practice (2nd Edition)*. London, UK: Sage

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H. and Earl, L. (2014) State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement* Vol 25, 2, 231-256

Murphy, R. (2013) *Testing Teachers: What works best for teacher evaluation and appraisal*. London, UK: Sutton Trust

NBPTS. (2001). *Early Adolescence through Young Adulthood / Art Standards*. Retrieved from <http://www.nbpts.org/standards-five-core-propositions/>

NBPTS. (2017). National Board for Professional Teaching Standards. Retrieved August 23, 2017, from <http://www.nbpts.org/standards-five-core-propositions/>

OECD. (2013). *Teachers for the 21st Century: Using Evaluation to Improve Teaching*. Retrieved from <http://link.springer.com/10.1007/978-3-642-36970-4>

OECD. (2014). *TALIS 2013 Results: An International Perspective on Teaching and Learning, TALIS*. Retrieved from <http://dx.doi.org/10.1787/9789264196261-en>

Papay, J. P. (2012). Refocusing the Debate: Assessing the Purposes and Tools of Teacher Evaluation. *Harvard Educational Review*, 82(1), 123–141.

Paulo, A. (2014) Pre-Service Teacher's Preparedness to Implement Competence-Based Curriculum in Secondary Schools in Tanzania. *International Journal of Education and Research* Vol 2, 7, 219-230

Peeraer, J., Kabanda, C., Nzabwirwa, W., Nizeyimana, G. and Uworwabayeho, A. (2015) Conceptions of learning and uptake of learner-centered pedagogy in initial teacher education in Rwanda, paper presented to UKFIET Conference – Learning for Sustainable Futures, University of Oxford, September 2015. Available

at: https://www.researchgate.net/profile/Jef_Peeraer/publication/284171026_Conceptions_of_learning_and_uptake_of_learner-centered_pedagogy_in_initial_teacher_education_in_Rwanda/links/564dc6be08ae1ef9296ace81.pdf (Accessed: 7 September 2017)

Popham, W. J. (1988) The Dysfunctional Marriage of Formative and Summative Teacher Evaluation. *Journal of Personnel Evaluation* 1, 269-273

Republic of Rwanda (2007) *Teacher Development and Management Policy in Rwanda*. Kigali, Rwanda: Ministry of Education

Republic of Rwanda (2013) *Education Sector Strategic Plan 2013/14-2017/8*. Kigali, Rwanda: Ministry of Education

Republic of Rwanda (2015) *Competence-Based Curriculum: Curriculum Framework – Pre-Primary to Upper Secondary*. Kigali, Rwanda: Ministry of Education – Rwanda Education Board (REB)

Republic of Rwanda (2016) *2016 Education Statistical Yearbook*. Kigali, Rwanda: Ministry of Education

Reynolds, D., Muijs, D. and Treharne, D. (2003) Teacher Evaluation and Teacher Effectiveness in the United Kingdom. *Journal of Personnel Evaluation in Education* Vol 17, 1, 83-100

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. and Stringfield, S. (2014) Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement*, Vol 25, 2, 197-230

Rutaisire, J. and Gahima, C. (2009) Policy and research in a post-conflict context: Issues and challenges in the implementation of the Rwandan Teacher Development and Management Policy. *International Journal of Educational Management*, Vol 23, 4, 326-335

SABER (2012) *What matters most in teacher policies? A framework for building a more effective teaching profession*. Washington, DC, USA: The World Bank

Sahlberg, P. (2006). Education reform for raising economic competitiveness. *Journal of Educational Change*, 7(4), 259–287. <http://doi.org/10.1007/s10833-005-4884-6>

Sahlberg, P. (2011). The Fourth Way of Finland. *Journal of Educational Change*, 12(2), 173–185. <http://doi.org/10.1007/s10833-011-9157-y>

Sammons, P. and Ko, J. (2008) *Effective Classroom Practice (ECP) Project Report Using Systematic Classroom Observation Schedules to Investigate Effective Teaching: Overview of Quantitative Findings*. School of Education University of Nottingham, August 2008

Santelices, M. V., Valencia, E., Gonzalez, J., & Taut, S. (2016). *Two teacher quality measures and the role of context: evidence from Chile*. *Educational Assessment, Evaluation and Accountability*. Educational Assessment, Evaluation and Accountability. <http://doi.org/10.1007/s11092-016-9247-8>

Santiago, P., & Benavides, F. (2009). *Teacher Evaluation: A Conceptual Framework and examples of Country Practices*.

Schmelkes, S. (2015). Assessment of teacher performance-state of affairs. In *Critical issues for formulating new teacher policies in Latin America and the Caribbean: the current debate*. UNESCO 2015.

Schultz, S., & Pecheone, R. (2014). Assessing Quality Teaching in Science. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems : new guidance from the measures of effective teaching project* (First). San Francisco, CA: Bill & Melinda Gates foundation | Jossey-Bass A Wiley Brand.

SEP. (2016). Etapas, Aspectos, Métodos e Instrumentos. Proceso de Evaluación del Desempeño Docente y Técnico Docente: Educación Básica. Secretaría de Educación Pública

SEP. (2017). *Perfil, parámetros e indicadores para docentes y técnicos docentes en educación básica*.

Sørensen, T. B. (2016). *Value-added measurement or modelling (VAM)*. University of Bristol.

SREE (no date) *History*. Available at <https://www.sree.org/>. Accessed: 22 September 2017

Stronge, J. H. (2010). *Handbook on the Teacher Performance Evaluation System*.

Taut, S., Santelices, V., Araya, C., & Manzi, J. (2010). Theory underlying a national teacher evaluation program. *Evaluation and Program Planning*, 33(4), 477–486.

<http://doi.org/10.1016/j.evalprogplan.2010.01.002>

Taut, S., Valencia, E., Palacios, D., Santelices, M. V, Jiménez, D., & Manzi, J. (2014). Teacher performance and student learning: linking evidence from two national assessment programmes. *Assessment in Education: Principles, Policy & Practice*, 23(September 2015), 1–22.

<http://doi.org/10.1080/0969594x.2014.961406>

Taylor, B. E. S., & Tyler, J. H. (2012). The Effect of Evaluation on Teacher Performance. *American Economic Review*, 102(7), 3628–3651.

Teddlie, C., & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer Pres.

Teddlie C., Stringfield, S. and Burdett, J. (2003) International Comparisons of the Relationships among Educational Effectiveness, Evaluation and Improvement Variables: An Overview. *Journal of Personnel Evaluation in Education* Vol 17, 1, 5-20

The Danielson Group (n.d.) *FFT Clusters - Math*. Available from www.danielsongroup.org/download/?download=1843 (Accessed: 22 September 2017)

The UTeach Institute. (n.d.). UTeach Observation Protocol. Retrieved September 27, 2017, from <https://utop.uteach.utexas.edu>

Tornero, B., & Taut, S. (2010). A mandatory, high-stakes National Teacher Evaluation System: Perceptions and attributions of teachers who actively refuse to participate. *Studies in Educational Evaluation*, 36(4), 132–142. <http://doi.org/10.1016/j.stueduc.2011.02.002>

Vázquez Cruz, M. del Á., Cordero Arroyo, G., & Leyva Barajas, Y. E. (2014). COMPARATIVE ANALYSIS OF PROFESSIONAL PERFORMANCE CRITERIA FOR TEACHING IN FOUR COUNTRIES OF AMERICA. *Actualidades Investigativas En Educación*, 14(3), 1–20

Villafuerte, L. P., Quinto Simón, S., Navarro Corona, C., Leyva Barajas, Y. E., & González Nava, M. F. (2015). La pertinencia del portafolio como instrumento de evaluación desde la perspectiva de los docentes. In *Segundo Congreso Latinoamericano de Medición y Evaluación Educacional*. Mexico City: colmee.

Wanzare, Z. O. (2002) Rethinking Teacher Evaluation in the Third World – the Case of Kenya, *Educational Management and Administration*, Vol. 30: 213-229

Westbrook J, Durrani N, Brown R, Orr D, Pryor J, Boddy J, Salvi F (2013) *Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries*. Final

Report. Education Rigorous Literature Review. London, UK: Department for International Development.

Whitley, M. J. *An exploration into whether the developmental appraisal system is achieving its desired outcome of promoting teacher development* (unpublished masters dissertation, University of Stellenbosch, 2016)

World Bank (2011) *Rwanda: Education Country Status Report – Toward Quality Enhancement and Achievement of Universal Nine Year Basic Education.* Washington, DC, USA: The World Bank

REPUBLIC OF RWANDA



RWANDA EDUCATION BOARD (REB)
P.O.BOX 3817 KIGALI

LESSON EVALUATION SHEET

School name: Teacher's name :.....
 Qualification :..... Experience :...years ;Class :.....Number of pupils :.....out of.....
 Subject :.....Topic :
 Period :.....Date :.....

No	Pedagogic aspects to be evaluated	Inspection scores					Comments
		5	4	3	2	1	
		A	B	C	D	E	
1.	The teacher's teaching documents (primary or secondary schools) are available, updated and regularly well prepared						
2.	Learner's documents are maintained and regularly checked						
3.	The lesson is effectively planned						
4.	The teacher masters the content of the lesson						
5.	Relevance and appropriateness of content						
6.	Methods used enable all the learners to learn effectively						
7.	Teacher manages learners well and insists on high standards of behavior						
8.	Evaluation of learners' work is effectively done						
9.	Learners are engaged in the lesson, sustain concentration and make significant progress						
10.	The use of the medium of instruction and communication skills						
11.	The management of time and instructional material is effective						

Overall performance:
 Excellent (≥47) ; Very good(≥38) ; Good(≥33) Satisfactory(≥27) ; Unsatisfactory<27)

Conclusion and feed back:
Strong points

Areas for improvement

Teacher's comments.

Inspector's name and signature _____ Teacher's name and signature _____

Appendix B – Example of a school framework based on the Teacher Standards for England

Formal Observation Checklist [CIRCLE...] **ACHIEVEMENT:** 1 2 3 4 **QUALITY OF TEACHING:** 1 2 3 4 **BEHAVIOUR AND SAFETY:** 1 2 3 4 **SMSC:** 1 2 3 4

NAME: _____ OBSERVER: _____ DATE: _____ CLASS: _____ PURPOSE: OVERALL JUDGEMENT:

If paired: _____

WWW /** from previous observation:

EBI/next step from previous observation:

Focus	Outstanding 1	Good 2	Requires improvement 3	Inadequate 4	Notes
1. Use of assessment in planning <i>(Teacher Standard 6)</i>	<input type="checkbox"/> Information from assessments is used to set tasks that are perfectly matched to the students' prior attainment. <input type="checkbox"/> Learning objectives are appropriate and allow all students to make rapid and sustained progress.	<input type="checkbox"/> Information from assessments is used to set tasks that are well matched to groups of students' prior attainment. <input type="checkbox"/> Learning objectives allow most students to make good progress and achieve well over time.	<input type="checkbox"/> Information from assessments is used to set tasks that are sometimes well matched to the needs of the class. <input type="checkbox"/> Learning objectives allow most students to make satisfactory progress.	<input type="checkbox"/> Information from assessments is not used in planning. <input type="checkbox"/> Learning objectives are inappropriate for the students resulting in little progress.	
2. Level of challenge <i>(Teacher Standards 1+2)</i>	<input type="checkbox"/> The work is pitched at a level that is achievable if individual students work hard and try their very best. <input type="checkbox"/> Teachers have consistently high expectations.	<input type="checkbox"/> The work is pitched at a level that is achievable for most within the group if they work hard and try their best. <input type="checkbox"/> Teachers have high expectations.	<input type="checkbox"/> The work is pitched at a level that is achievable for all within the class with extension activities and support mechanisms in place. <input type="checkbox"/> Some students find the work too easy or too hard.	<input type="checkbox"/> Tasks are not suitably matched to students' prior attainment. <input type="checkbox"/> Students find the work too easy or too hard. <input type="checkbox"/> Teachers do not have sufficiently high expectations.	

			<input type="checkbox"/> Teachers do not have sufficiently high expectations.		
3.Support for students with specific learning needs (Teacher Standard 5)	<input type="checkbox"/> Students with specific learning needs receive support at the time and level it is required to optimise their learning.	<input type="checkbox"/> Students with specific learning needs receive support that allows them to work independently and make good progress during the lesson.	<input type="checkbox"/> Students with specific learning needs receive sporadic support that allows most of them to make satisfactory progress.	<input type="checkbox"/> Students with specific learning needs are not provided with appropriate support and guidance to allow them to make progress.	

Focus	Outstanding 1	Good 2	Requires improvement 3	Inadequate 4	Notes
4.Assessment for learning (Teacher Standard 3)	<input type="checkbox"/> Questions tease out students' understanding. <input type="checkbox"/> Marking is frequent and regular, providing students with very clear guidance on how work can be improved. <input type="checkbox"/> Teachers systematically and effectively check students understanding, intervening with notable impact.	<input type="checkbox"/> Questions tease out most students' understanding and tasks are reshaped if necessary to improve learning. <input type="checkbox"/> Marking is frequent and regular, providing students with guidance on how work can be improved. <input type="checkbox"/> Teachers check students understanding.	<input type="checkbox"/> Questions are not used effectively to assess understanding. <input type="checkbox"/> Students have to wait with their hands up for support. <input type="checkbox"/> Marking is encouraging but provides little guidance on how work can be improved.	<input type="checkbox"/> Questions are closed and are not used to assess students' understanding. <input type="checkbox"/> Marking is infrequent and/or irregular and/or fails to provide the pupil with guidance on how work can be improved.	
5.Use of resources (Teacher Standard 4)	<input type="checkbox"/> Resources, including teaching assistants where applicable, are highly effective in prompting rapid learning for groups of students regardless of their aptitudes and needs. <input type="checkbox"/> Homework is an integral part of the lesson. It extends the learning and is treated as being as important as the lesson itself.	<input type="checkbox"/> Resources, including teaching assistants where applicable, are well deployed to support learning for groups of students regardless of their aptitudes and needs so that all such groups make at least good progress. <input type="checkbox"/> Homework is used to extend the learning.	<input type="checkbox"/> Resources, including teaching assistants where applicable, sometimes provide suitable support for groups of students regardless of their aptitudes and needs. <input type="checkbox"/> Homework is not used to extend the learning.	<input type="checkbox"/> Resources, including teaching assistants where applicable, do not meet the needs of groups of students so that their learning is limited by too much being done for them. <input type="checkbox"/> Homework is not set or is seen as a 'bolt on' with little relevance to the lesson.	

6. Opportunities to develop RWCM and competencies	<input type="checkbox"/> The work includes good opportunities to develop students' skills in RWCM. <input type="checkbox"/> The lesson provides opportunities for extending wider skills. <input type="checkbox"/> Students are given opportunities to lead their own learning.	<input type="checkbox"/> The work includes some opportunities to develop students' skills in RWCM. <input type="checkbox"/> Competency skills are developed. <input type="checkbox"/> Students are given opportunities to work independently.	<input type="checkbox"/> The work includes very few opportunities to develop students' RWCM and competency skills.	<input type="checkbox"/> The promotion of basic skills is ignored.	
----------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------	--

Focus	Outstanding 1	Good 2	Requires improvement 3	Inadequate 4	Notes
7. Level of engagement <i>(Teacher Standard 7)</i>	<input type="checkbox"/> The tasks themselves enthuse students so that they persevere when challenged and are keen to succeed and to learn more. <input type="checkbox"/> They are fully engaged on tasks throughout the lesson and committed to learning.	<input type="checkbox"/> Students find the tasks interesting and enjoyable. <input type="checkbox"/> They concentrate well and pay full attention to the teacher.	<input type="checkbox"/> Some students are not interested in the tasks and are disengaged leading to off task behaviour.	<input type="checkbox"/> Students are bored by the tasks , and become disruptive.	
8. Behaviour for learning <i>(Teacher Standard 7)</i>	<input type="checkbox"/> Students are engrossed in their work. <input type="checkbox"/> There is no need for any overt discipline. <input type="checkbox"/> Teachers have consistently high expectations. <input type="checkbox"/> There is a classroom climate of shared endeavour.	<input type="checkbox"/> Students are engaged in their work. <input type="checkbox"/> Only the briefest of reminders is required to refocus students. <input type="checkbox"/> Teachers create a positive climate for learning.	<input type="checkbox"/> Some students are off task and display low level disruptive behaviour which has an effect on their progress. <input type="checkbox"/> Too much time is spent managing behaviour.	<input type="checkbox"/> Most students are not engaged with the tasks. Poor behaviour impacts negatively and at the expense of learning.	
9. SMSC	Positive role modelling e.g. of equalities, diversity. Fostering independence, love of subject, positive peer relationships, understanding of right and wrong. Provides opportunities for wider debate about issues, reflection etc.				



Tackles inappropriate language / behaviour effectively.

WWW/ **:

SIGNED TEACHER:

EBI/next steps:

SIGNED OBSERVER:

Appendix C – Quality Science Teaching (QST) and QST-MET rubrics

Domain 1: Assessing Teacher's Content Knowledge and Pedagogy

Indicators	1 Provides little or no evidence	2 Provides limited evidence	3 Provides clear evidence	4 Provides consistently strong evidence
I.1. Sets the context and focuses learning on key science concepts	The teacher does <i>not</i> set the context for the lesson AND does <i>not</i> provide objectives/goals OR focuses on loosely connected biological facts and definitions.	The teacher sets the context for the lesson in general terms and provides vague objectives/goals OR occasionally focuses on relevant biological concepts.	The teacher sets the context for the lesson in clear terms and provides clear objectives/goals OR focuses learning primarily on relevant biological concepts.	The teacher explicitly sets the context for the lesson in clear and specific terms and provides clear and worthwhile objectives/goals OR continually promotes scientific thinking and coherent conceptual understanding.
I.2. Uses representations Representations include examples, analogies, models, simulations	The teacher does <i>not</i> use representations OR uses representation(s) that lead students to misunderstand or misinterpret key biological concepts.	The teacher uses representation(s) that are general in nature OR that may limit the students' ability to understand or distinguish among related biological concepts.	The teacher uses multiple representations that are similar in nature and enhance students' ability to understand or distinguish among related biological concepts.	The teacher uses multiple representations that are different but related, enabling students to understand and distinguish among key biological concepts and ways of thinking/doing biology.
I.3. Demonstrates content knowledge	The teacher makes significant and/or numerous inaccuracies about main biological concepts and processes OR the teacher does <i>not</i> recognize or does <i>not</i> address students' preconceptions.	The teacher makes minor OR occasional inaccuracies but does <i>not</i> immediately correct the error about the main biological concepts and processes OR the teacher recognizes students' preconception and indicates the student is incorrect.	The teacher makes minor OR occasional inaccuracies AND immediately corrects the error related to biological concepts and processes OR the teacher recognizes and corrects students' preconceptions continuing with the planned lesson.	The teacher makes no observed inaccuracies related to biological concepts and processes AND the teacher recognizes and utilizes students' preconceptions as an entry point for students to engage with key biological concepts.

Cluster 1: Assessing Teacher’s Knowledge and Pedagogy

Indicators	1 Provides little or no evidence	2 Provides limited evidence	3 Provides clear evidence	4 Provides consistently strong evidence
1. Sets the context and focuses learning on key science concepts	The teacher does <i>not</i> set the context for the lesson AND does <i>not</i> provide objectives/goals OR focuses on loosely connected biological facts and definitions.	The teacher sets the context for the lesson in general terms and provides vague objectives/goals OR occasionally focuses on relevant biological concepts.	The teacher sets the context for the lesson in clear terms and provides clear objectives/goals OR focuses learning primarily on relevant biological concepts.	The teacher explicitly sets the context for the lesson in clear and specific terms and provides clear and worthwhile objectives/goals, AND continually promotes scientific thinking and coherent conceptual understanding.
2. Uses representations. Representations include examples, analogies, models, simulations	The teacher does <i>not</i> use representations OR uses representation(s) that lead students to misunderstand or misinterpret key biological concepts.	The teacher uses representation(s) that are general in nature OR that may limit the students’ ability to understand or distinguish among related biological concepts.	The teacher uses multiple representations that are similar in nature and enhance students’ ability to understand or distinguish among related biological concepts.	The teacher uses multiple representations that are different but related, enabling students to understand and distinguish among key biological concepts and ways of thinking/doing biology.
3. Demonstrates content knowledge	The teacher makes significant and/or numerous inaccuracies about main biological concepts and processes OR the teacher does <i>not</i> recognize or does <i>not</i> address students’ preconceptions.	The teacher makes minor OR occasional inaccuracies but does <i>not</i> immediately correct the error about the main biological concepts and processes, OR the teacher recognizes students’ preconception and indicates the student is incorrect.	The teacher makes minor OR occasional inaccuracies AND immediately corrects the error related to biological concepts and processes, OR the teacher recognizes and corrects students’ preconceptions continuing with the planned lesson.	The teacher makes no observed inaccuracies related to biological concepts and processes, AND the teacher recognizes and utilizes students’ preconceptions as an entry point for students to engage with key biological concepts.
4. Provides students with feedback	The teacher provides no or nonspecific feedback to students/class.	The teacher provides specific feedback to students/class that focuses on the general strengths of the work OR how to improve learning.	The teacher provides specific feedback to students/class that focuses on the general strengths of the work AND how to improve their learning.	The teacher provides constructive, specific, detailed feedback to students/class AND includes explicit suggestions on how to improve their learning.

Appendix D

UTeach Observation Protocol for Mathematics and Science

Complete **AFTER** observation of lesson using field notes, teacher post-interview, and student work samples and/or comments (video if available).⁵

I. BACKGROUND INFORMATION

Teacher: [Type text]

School: [Type text]

Date of Observation: [Type text]

Start and End Time of Observation: [Type text]

Date of Post Interview: [Type text]

Subject Observed: [Type text]

Grade Level: [Type text]

Course Level: (Regular or Advanced/Accelerated): [Type text]

Observer: [Type text]

II. LESSON OVERVIEW

Lesson Description

In a paragraph or two, describe the lesson you observed. Include where the lesson fits into the overall unit of study. Be sure to include enough detail to provide a context for your ratings of the lesson and also to allow you to recall the details of the lesson when needed in the future.

III. RATING SCALES

1 = Not observed at all / Not demonstrated at all	4 = Observed often / Demonstrated well
2 = Observed rarely / Demonstrated poorly	5 = Observed to a great extent / Demonstrated to a great extent
3 = Observed an adequate amount / Demonstrated adequately	

Note About Synthesis Ratings

The synthesis ratings are **not intended to be a mathematical average of the indicator scores** making up each section, but are designed to allow the rater to describe his or her overall impression, using a holistic view of the domain and providing a “human average” of the entire lesson. Evidence to support the score chosen can be typed in the open space after the Synthesis Ratings boxes.

⁵ NOTE: The UTOP was adapted from Horizon Research, Inc., *2005–06 Core Evaluation Manual: Classroom Observation Protocol* by UTeach Natural Sciences, University of Texas at Austin.

1. Classroom Environment

Rating	Indicator
	<p>1.1 Classroom Engagement: The classroom environment facilitated by the teacher encouraged students to generate ideas, questions, conjectures, and/or propositions that reflected engagement or exploration with important mathematics and science concepts.</p> <p>Description, Rubric, and Examples</p>
	<p>1.2 Classroom Interactions: Interactions reflected collegial working relationships among students (e.g., students worked together productively and talked with each other about the lesson).</p> <p><i>*It's possible that this indicator was not applicable to the observed lesson. You may rate NA in this case.</i></p> <p>Description, Rubric, and Examples</p>
	<p>1.3 Classroom On-Task: The majority of students were on task throughout the class.</p> <p>Description, Rubric, and Examples</p>
	<p>1.4 Classroom Management: The teacher's classroom management strategies enhanced the classroom environment.</p> <p>Description, Rubric, and Examples</p>
	<p>1.5 Classroom Organization: The classroom is organized appropriately such that students can work in groups easily and get to lab materials as needed, and the teacher can move to each student or student group.</p> <p>Description, Rubric, and Examples</p>
	<p>1.6 Classroom Equity: The classroom environment established by the teacher reflected attention to issues of access, equity, and diversity for students (e.g., cooperative learning, language-appropriate strategies and materials, attentiveness to student needs).</p> <p>Description, Rubric, and Examples</p>

Synthesis Rating for Classroom Environment

Classroom culture is <i>non-interactive</i> or <i>non-productive</i> .	Classroom culture is productive and interactive only <i>occasionally</i> .	Classroom culture is <i>adequately</i> productive and interactive.	Classroom culture is <i>often</i> productive and interactive, with some collegial interactions.	Classroom culture is <i>consistently</i> collegial, interactive, and productive.
1	2	3	4	5

2. Lesson Structure

Rating	Indicator
	<p>2.1 Lesson Sequence: The lesson was well organized and structured (e.g., the objectives of the lesson were clear to students, and the sequence of the lesson was structured to build understanding and maintain a sense of purpose).</p> <p>Description, Rubric, and Examples</p>
	<p>2.2 Lesson Importance: The structure of the lesson allowed students to engage with and/or explore important concepts in mathematics or science (instead of focusing on techniques that may only be useful on exams).</p> <p>Description, Rubric, and Examples</p>
	<p>2.3 Lesson Assessments: The structure of the lesson included opportunities for the instructor to gauge student understanding.</p> <p>Description, Rubric, and Examples</p>
	<p>2.4 Lesson Investigation: The lesson included an investigative or problem-based approach to important concepts in mathematics or science.</p> <p>Description, Rubric, and Examples</p>
	<p>2.5 Lesson Resources: The teacher obtained and employed resources appropriate for the lesson.</p> <p>Description, Rubric, and Examples</p>
	<p>2.6 Lesson Reflection: The teacher was critical and reflective about his/her practice after the lesson, recognizing the strengths and weaknesses of his/her instruction.</p> <p><i>* This indicator may be rated NA if you do not have access to a teacher interview or teacher commentary.</i></p> <p>Description, Rubric, and Examples</p>

Synthesis Rating for Lesson Structure

Lesson was <i>very poorly</i> structured to assist student learning.	Lesson was <i>poorly</i> structured to assist student learning.	Lesson was <i>adequately</i> structured to assist student learning.	Lesson was <i>well</i> structured to assist student learning.	Lesson was <i>expertly</i> structured to assist student learning.
1	2	3	4	5

3. Implementation

Rating	Indicator
	<p>3.1 Implementation Questioning: The teacher used questioning strategies to encourage participation, check on skill development, and facilitate intellectual engagement and productive interaction with students about important science and mathematics content and concepts.</p> <p>Description, Rubric, and Examples</p>
	<p>3.2 Implementation Involvement: The teacher involved all students in the lesson (calling on non-volunteers, facilitating student–student interaction, checking in with hesitant learners, etc.).</p> <p>Description, Rubric, and Examples</p>
	<p>3.3 Implementation Modification: The teacher used formative assessment effectively to be aware of the progress of all students and modified the lesson appropriately when formative assessment demonstrated that students did not understand.</p> <p>Description, Rubric, and Examples</p>
	<p>3.4 Implementation Timing: An appropriate amount of time was devoted to each part of the lesson.</p> <p>Description, Rubric, and Examples</p>
	<p>3.5 Implementation Connections: The instructional strategies and activities used in this lesson clearly connected to students’ prior knowledge and experience.</p> <p>Description, Rubric, and Examples</p>
	<p>3.6 Implementation Safety: The teacher’s instructional strategies included safe, environmentally appropriate, and ethical implementation of laboratory procedures and/or classroom activities.</p> <p><i>*This indicator may be rated NA if there were no relevant activities during the lesson.</i></p> <p>Description, Rubric, and Examples</p>

Synthesis Rating for Implementation

<i>Very poor</i> lesson implementation	<i>Poor</i> lesson implementation	<i>Adequate</i> lesson implementation	<i>Good</i> lesson implementation	<i>Excellent</i> lesson implementation
1	2	3	4	5

4. Mathematics/Science Content

Rating	Indicator
	<p>4.1 Content Significance: The mathematics or science content chosen was significant, worthwhile, and developmentally appropriate for this course (includes the content standards covered, as well as examples and activities chosen by the teacher).</p> <p>Description, Rubric, and Examples</p>
	<p>4.2 Content Fluency: Content communicated through direct and non-direct instruction by the teacher is consistent with deep knowledge and fluency with the mathematics or science concepts of the lesson (e.g., fluent use of examples, discussions, and explanations of concepts, etc.).</p> <p>Description, Rubric, and Examples</p>
	<p>4.3 Content Accuracy: Teacher written and verbal content information was accurate.</p> <p>Description, Rubric, and Examples</p>
	<p>4.4 Content Assessments: Formal assessments used by teacher (if available) were consistent with content objectives (homework, lab sheets, tests, quizzes, etc.).</p> <p><i>*It's possible that this indicator was not applicable to the observed lesson. You may rate NA in this case.</i></p> <p>Description, Rubric, and Examples</p>
	<p>4.5 Content Abstraction: Elements of mathematical/scientific abstraction were used appropriately (e.g., multiple forms of representation in science and mathematics classes include verbal, graphic, symbolic, visualizations, simulations, models of systems and structures that are not directly observable in real time or by the naked eye, etc.).</p> <p><i>*It's possible that this indicator was not applicable to the observed lesson. You may rate NA in this case.</i></p> <p>Description, Rubric, and Examples</p>
	<p>4.6 Content Relevance: During the lesson, it was made explicit to students why the content is important to learn.</p> <p>Description, Rubric, and Examples</p>
	<p>4.7 Content Interconnections: Appropriate connections were made to other areas of mathematics or science and/or to other disciplines.</p> <p>Description, Rubric, and Examples</p>
	<p>4.8 Content Societal Impact: During the lesson, there was discussion about the content topic's role in history, current events, or relevant "real-world" problems.</p> <p>Description, Rubric, and Examples</p>

Synthesis Rating for Content

Students learning <i>inaccurate</i> content knowledge	Students learning <i>superficial</i> content knowledge	Students learning <i>adequate</i> content knowledge	Students learning <i>good</i> content knowledge	Students learning <i>deep, fluid</i> content knowledge
1	2	3	4	5

IV. SUMMARY COMMENTS

Information included in this section is a snapshot of your evaluation of the quality of the lesson. When filling in this section, consider all available information concerning the lesson and its context and purpose, as well as your own judgment of the relative importance of the ratings given. The summary is intended to be freeform and can also include comments that did not fit into any of the preceding sections.

FIELD NOTES

Use this space to take field notes, capture comments from student–student or student–teacher conversations, describe the physical, socio-emotional, or cultural environment of the classroom interactions, and so on. Field notes can be edited and inserted into the Evidence boxes under each indicator to illustrate your rationale for assigning a particular score for that indicator.

Be sure to REMOVE all notes prior to sharing with anyone!