Understanding the classroom

A guide to selecting classroom observation tools

Jenny Price, Katarzyna Kubacka, Geeta Gambhir,
Maria-Jose Guevara, Sébastien Hine

NFER
International

Acknowledgements

All learners have a right to quality, inclusive education. Understanding what happens in classrooms, while taking into account the wider societal context, is a fundamental step towards this goal.

This guide is an introduction to classroom observation tools and gives an overview and comparison of three commonly used tools – Stallings, CLASS and Teach.

Intended for use by various audiences, such as practitioners, school stakeholders and policy makers in low- and middle-income countries, this guide aims to help identify the most suitable tool for their particular needs.

At a glance

Stallings

- For primary and secondary
- Four languages
- Best for capturing basic information on teaching practices

CLASS

- For early childhood through secondary
- Three languages
- Best for early childhood phase and capturing richer data

Teach

- For primary but early childhood and secondary under development
- Twelve languages
- Best for multi-languages and collecting low and high inference data

When to use classroom observation tools

There are a number of factors that impact teaching and learning in classrooms, ranging from global issues like the COVID 19 pandemic, to cultural norms and educational policy, right down to more personal issues such as teacher motivation and attitudes towards equality.

The tools discussed have the potential to offer important insights into how classroom practices may be improved to the benefit of all learners.

This guide focuses on three main educational issues, described below: teachers' classroom practices; supporting students' inclusion; and holistic learning environments.

Teachers' classroom practices – looking at how teachers teach. Classroom observation systems can help answer questions regarding teachers' motivations, skills, behaviour management, student engagement, and the support and resources needed for success.

Supporting students' inclusion – all students should have access to quality education. In order to tackle the disproportionate numbers of girls and children with disabilities still excluded from learning, understanding classroom practices can support teacher development and identify the need for specialist training.

Holistic learning environments – socio-emotional learning, those skills beyond literacy and numeracy, are increasingly recognised to be beneficial to students' long-term success. Observing how these skills are nurtured and teacher-student interactions, can help inform schools' practices and wider national curriculums.

Things to be mindful of

It is important to choose the tool best suited to your particular context. Each tool has certain limitations including language, required resource and range of data that could affect the findings.

Similarly, the tools are undergoing constant development, often addressing elements of these limitations, so it is advised to check the relevant resources for the most recent iteration.

1 Conducting Classroom Observations: What? Who? Why?

1.1 Introduction to this guide: What is its purpose, and who is it for?

Sustainable Development Goal 4 calls for global efforts to achieve quality and inclusive education for all, and acts as a mobilising force for governments, civil society organisations, communities and practitioners. To reach this goal, we need to understand the starting point of the journey – the teaching and learning as it happens in the classroom.

Understanding what actually happens inside classrooms is particularly relevant in low and middle income countries (LMICs), where education systems face a range of context driven challenges and barriers to achieving SDG4. As a result, classroom and teacher observation tools have gained increasing prominence on the international education and development agenda as a way to better understand how teaching practices impact student learning.

Written for researchers, practitioners, evaluators, and other education stakeholders, this guide:

- Introduces observational methods to generate evidence of what is happening in the classroom

- Summarises the relative advantages and disadvantages of three widely used tools

- Provides insights for those looking to identify the most appropriate tool for their needs

- Signposts further information for those wanting to learn more about observational methods

We summarise the existing literature relating to three classroom observation tools commonly used to explore teaching practices in LMIC contexts: The Stallings Classroom Observation System (Stallings); The Classroom Assessment Scoring System (CLASS); and Teach. This review was supplemented with 15 interviews with experts who have extensive experience developing, evaluating, and implementing classroom observation tools.

The guide responds to three core research questions:

1. What are classroom observation tools and how are they used in international contexts?

2. What are the advantages and shortcomings of widely used classroom observation tools?

3. How do Stallings, CLASS and Teach tools relate to key international policy priorities on gender, disability and socio-emotional learning?

1.2 Why classroom observation matters?

Broader cultural, social and economic factors and individual characteristics all affect the interactions between students and teachers. Hence, in trying to understand teaching and learning, it is important to consider the wider societal context, as well as more personal factors, which influence these processes. Classroom observation can shed light on these characteristics and help unpack their influence on learning.

To illustrate this, we consider three educational issues:

1: Teachers' classroom practices

2: Supporting students' inclusion in the classroom
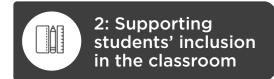

3: Holistic learning environments

1.2.1 Issue 1: Teachers' classroom practices

Generating evidence to understand how teachers teach is perhaps the most prominent reason for choosing observational methods. They allow the user to explore questions about what is actually happening during lessons, such as: How much time is spent actively teaching? Is the teacher able to manage student behaviour? What methods and techniques are teachers using to engage and motivate students? Are students actively engaged in learning?

Answering such questions can help capture the quality of teaching practices, to provide teachers with feedback for improvement, and to identify where additional support, resources or funds are required to improve education systems.

At the same time, teaching does not happen in a vacuum. There are a multitude of factors which might influence whether teachers have the skills, knowledge and motivation to teach effectively:

Low numbers of subject matter experts in specialist topics such as mathematics, science and technology

Poor teacher working conditions and low salaries leading to low motivation, absenteeism, lack of preparedness and high teacher turnover

Limited/no access to specialised resources or teacher training to support students with diverse needs and abilities

Factors which influence teachers skills, knowledge and motivations

Low supply of qualified and trained teachers

Reliance on rote learning techniques which fail to advance comprehension

Overcrowded classrooms and limited educational resources

These challenges are magnified by the intensive pressure on education systems in contexts where basic education is oversubscribed and underfunded. As a result, they further exacerbate the challenges facing teachers and students.

2: Supporting students' inclusion in the classroom

1.2.2 Issue 2: supporting students' inclusion in education

Achieving quality education and learning for all, regardless of need or ability, is at the heart of SDG4. Understanding if and how students are excluded or marginalised from the learning process within the classroom, and what works to overcome this, is a major global education priority. Existing global evidence indicates that while educational access has expanded significantly, certain groups remain disproportionately excluded from quality education opportunities.

Despite major improvements in gender parity, girls continue to face exclusion, particularly when gender intersects with other forms of marginalisation. Out of all children who have never received a formal education, three quarters are estimated to be girls. Children with disabilities are even more likely to be out of school, and less likely to learn when in school. Approximately 40 percent of children with disabilities at primary level age are not in school, which increases to 55 percent at secondary level[1].

There are many factors that impact whether girls and children with disabilities are able to *access* schools and challenges persist that prevent them from *learning effectively* while inside the classroom. The barriers faced by girls and children with disabilities are multidimentional.

- **Systemic factors** – such as lack of trained teachers to support students with special needs and to take gender-sensitive approaches.

- **Societal barriers** – such as prejudice and bias which fuel negative attitudes towards girls' education and inclusion or gendered expectations regarding girls' and boys' abilities, participation in the classroom or their choice of subjects.

In recognition of these ongoing challenges to inclusion, many education policies, programmes and interventions focused on teacher professional development and training increasingly include dimensions of gender and disability sensitive pedagogy. Observing classroom practices as they happen can give an important insight into the extent to which teachers are able (and willing) to include all students in the learning, and to identify where additional training, support or resources are required. It can also provide teachers with direct and timely feedback to improve their inclusive practices in real time.

3: Holistic learning environments

1.2.3 Issue 3: Building holistic learning environments through socio-emotional learning

A final focus of this paper is the extent to which observing teaching practices and classroom environments can tell us about the broader range of skills children gain beyond foundational skills such as literacy and numeracy. Various terminology such as 21st century skills, life skills, socio-emotional learning, non-cognitive skills, soft skills, and character or personality education is used to describe a broad range of non academic skills. This guide will use the term socio-emotional learning (SEL) to describe learning which emphasises growth and progress.

These aspects of learning have been shown to be important for students' life outcomes. For instance, the extent to which students are able to exhibit self-control or perseverance is related to their success at work but also to their health. These skills are also an important element of an environment which supports all students' learning and makes them feel safe and comfortable. This is especially important for disadvantaged students, such as those living in unstable households or at risk of violence, as they enter schools from stressful environments. Thus, they need safety and socio-emotional skill support to be ready to start the learning process[2].

There is growing attention from policy-makers and global influencers to understand how SEL is taught and nurtured in classroom settings, either through direct instruction using a specific curriculum; by altering the school environment, through teachers' practices and their style of interaction with students; by changing school rules and expectations; or by influencing students' mindsets.

1   UNESCO, 2020
2   Sanchez, 2021

2.1 How can classroom observations support teaching and learning?

Reflecting on the issues and challenges noted in Section 1, classroom observations can provide valuable insights into what is happening inside classrooms, and how this relates to student learning. Depending on the tool or approach used, classroom observations can serve different purposes for different audiences across each level of the education system, from individual teachers up to national and global policy makers. Some uses of classroom observations include:

- **Teacher appraisal** - to generate tailored feedback for individual teachers as part of continuous professional development initiatives. For instance, this can help improve teaching practices by providing teachers with formative feedback based on their real-life classroom behaviour.

- **Programme or policy evaluations** – to measure the effectiveness and impact of teacher training and classroom interventions. For instance, classroom observation tools can be used to investigate teacher practice before, during and after an intervention, such as administering a new teacher professional development programme, in order to better understand how training changes teaching.

- **Education system diagnostics** - results from analyses based on classroom observation tools, when summarised at school, regional or national level can add to the evidence base on resource needs and challenges and serve as input to inform policy change at school, regional or national level. For instance, classroom observation tools can reveal that teachers in very large classrooms struggle to spend enough time on instruction, which can be used to advocate for classroom size changes and additional resources.

Figure 1. Potential users of observation tools

Classroom, school or programme level

District, national, education system or global level

Users

- School principals
- Teacher managers
- Other teachers
- Ministries of education
- (I)NGOs
- UN agencies

- Ministries of education
- Teacher training colleges
- Development Banks
- UN agencies
- Think tanks
- Research institutes
- Universities

A note on 'low inference' versus 'high inference' tools

Tools are commonly referred to as low-inference or high-inference; inference being the level of subjective judgement required when making observations[3]:

- Low-inference tools: collect data on interactions between teachers and students, concentrating on aspects of the classroom that are directly observable and easy-to-code in formats such as checklists or binary yes/no responses. This generates robust comparable quantitative data on issues such as time-on-task, teachers' use of time, or teachers' use of different learning activities[4]. Low-inference tools typically collect simple quantitative data.

- High-inference tools: are more subjective and require observers to make judgements about the quality of interactions between teachers and students, often with rating scales instead of binary responses[5]. Examples include the extent to which teachers are using feedback loops, using open-ended questions, encouraging participation, or using encouragement and affirmation. High-inference tools typically capture richer qualitative data than low-inference tools

3; 4; 5   RTI International, 2016

2.2 The Stallings Classroom Observation System

Covers primary and secondary education	Low inference and high inter-rater reliability	Generates robust quantitative data	Open source materials in multiple languages

What is the Stallings classroom observation system, and what does it measure?

Developed in the 1970s by professor Jane Stallings, the Stallings Classroom Observation System (referred to hereafter as Stallings) was developed to research the efficiency and quality of teaching in basic education in the United States[6]. The tool comprises a questionnaire and protocol for timed observations, which collect quantitative 'snapshots' at evenly spaced intervals. This provides robust data, which is particularly useful in presenting a clear picture of what is happening in the classroom at the school, regional or system level[7].

The tool covers four variables:

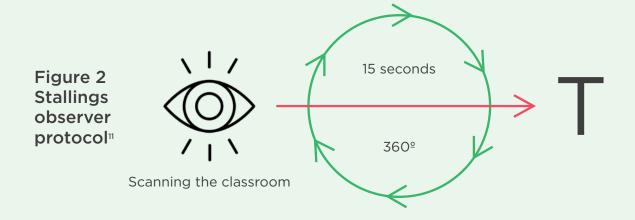- Teachers' use of instructional time
- Teachers' use of materials, including Information Communication Technology
- Core pedagogical practices
- Teachers' ability to keep students engaged

How does the tool work?

The Stallings observer protocol describes a structured approach to observing classroom activities. Ten discrete observations are made over the duration of each class at regular, evenly spaced intervals. The observer conducts a 15 second 'scan' of the room to capture teacher activities and student engagement. After each scan the observer selects from a pre-populated list of options to record:

- What the teacher is doing
- The material the teacher is using for the activity
- How many students are engaged in the activity
- What students not engaged in the activity are doing

The Stallings tool can produce comparable data across different schools and country contexts, and results can be measured against benchmark targets for various activities[8]. Figure 2 below sets out the proposed benchmarks for best practice use for instructional time, based on research conducted on effective teaching in the United States (US)[9]. It is recommended that these benchmarks are used as a reference point only - although grounded in several decades of research, they are based on the US context only.[10]

Figure 2 Stallings observer protocol[11]

Scanning the classroom

15 seconds

360º

6   World Bank, 2017
7   World Bank, 2017
8   Bruns and Luque, 2015; World Bank, 2017
9   Stallings and Knight, 2003
10   Bruns and Luque, 2015
11   World Bank, 2015

Table 1.1 Stallings use of time benchmark targets

	Instruction		Stallings benchmarks for effective time use = 85%
Teachers	Active instruction	Reading aloud	50% or more
		Demonstration/lecture	
		Discussion/question and answer	
		Practice and drill	
		Cooperative projects	
	Passive instruction	Monitoring copying	35% or less
		Monitoring seatwork	
	Classroom management		**15% or less**
	Verbal instruction (e.g., giving assignments)		
	Managing with students (e.g., passing out papers)		
	Disciplining students		
	Managing alone (e.g., grading homework)		
	Off-task		**0%**
	Absent from the room		
	Socialising with students		
	Socialising with others		
Students	**Off-task**		**6% or less**
	Socialising		
	Uninvolved		
	Acting out		

2.3 The Classroom Assessment Scoring System (CLASS)

Covers early childhood through to secondary level

High inference /in-depth measures

Captures teacher-student interactions

Accredited training required

What is CLASS, and what does it measure?

CLASS (Classroom Assessment Scoring System) was initially developed in the USA, and has since been implemented in over 50 countries[12]. Originally, CLASS focused on early years contexts only, but has now expanded to include six different protocols covering early childhood through to secondary level[13]. It is a high-inference tool that focuses on the teacher-child interactions that can improve the wellbeing and educational outcomes of children, and provides teachers with support and feedback to improve their instruction[14].

The tool explores three broad 'domains' that are divided into 11 'dimensions' which define the aspects of quality being measured, and each dimension is measured against a set of defined indicators. An additional 12th dimension of 'student engagement' captures the focus and participation of students in learning activities. The three domains are briefly described as:

- **Emotional Support**: considers the emotional connections and relationships between students and teachers; the responsiveness of teachers to the social/emotional needs of students; and the extent to which students are given autonomy and leadership opportunities, and that their ideas and opinions are respected and valued.

- **Classroom Management**: measures behaviour management and productivity, including the overall level of negativity among teachers and students; the methods teachers use to encourage desirable behaviour or redirect misbehaviour; and how well the teacher is able to manage time and routines to maximise the instructional time.

- **Instructional support**: assesses the instructional learning formats used to maximise student engagement; depth of lesson content and approaches to ensure understanding; facilitation of higher level thinking skills such as analysis and problem solving; quality of feedback provided to expand learning and encourage participation; and the degree to which teachers prompt and guide students to achieve a deeper understanding of content[15].

12   Teachstone n.d.a
13   Teachstone, n.d.b; Ofsted, 2018
14   Teachstone, n.d.c
15   Bruns *et al*, 2016.

How does the tool work?

CLASS is relatively complex to deliver, and requires researchers to be fully trained and certified to increase the reliability and validity of results. Researchers select timed segments of a class to apply the tool (typically between 15-25 minutes per segment), and use the CLASS guide and rubric to identify specific behaviours across each dimension. Each behaviour or dimension is scored as low, medium or high using a Likert scale, with low being a score of 1–2, medium a score of 3–5, and high a score of 6–7[16]. Classes can also be videotaped and reviewed later by a qualified observer or evaluator[17]. The below table provides an example of the rubric for scoring one dimension (behaviour management).

Behaviour Management			
	Low (1,2)	Medium (3,4,5)	High (6,7)
Clear Behaviour Expectations • Clear expectations • Consistency • Clarity of rules	Rules and expectations are absent, unclear, or inconsistently enforced.	Rules and expectations may be stated clearly, but are inconsistently enforced.	Rules and expectations for behaviour are clear and are consistently enforced.
Proactive • Anticipates problem behaviour or escalation • Rarely reactive • Monitoring	Teacher is reactive and monitoring is absent or ineffective.	Teacher uses a mix of proactive and reactive responses; sometimes monitors but at other times misses early indicators of problems.	Teacher is consistently proactive and monitors effectively to prevent problems from developing.
Redirection of Misbehaviour • Effectively reduces misbehaviour • Attention to the positive • Uses subtle cues to redirect • Efficient	Attempts to redirect misbehaviour are ineffective; teacher rarely focuses on positives or uses subtle cues. As a result, misbehaviour continues/escalates and takes time away from learning.	Some attempts to redirect misbehaviour are effective; teacher sometimes focuses on positives and uses subtle cues. As a result, there are few times when misbehaviour continues/escalates or takes time away from learning.	Teacher effectively redirects misbehaviour by focusing on positives and making use of subtle cues. Behaviour management does not take time away from learning.
Student Behaviour • Frequent compliance • Little aggression and defiance	There are frequent instances of misbehaviour in the classroom.	There are periodic episodes of misbehaviour in the classroom.	There are few, if any, instances of student misbehaviour in the classroom.

16   MET Project, 2010; Bruns *et al*, 2016
17   World Bank, 2017

2.4 Teach

| Covers ECE to Primary level | High and low inference data collected | Designed specifically for LMICs | Open source tools in multiple languages |

What is Teach, and what does it measure?

Teach was launched in 2019 by the World Bank in response to the demand for an open source classroom observation tool designed specifically for use in LMICs. It was launched as part of the World Bank's new strategic approach of "Successful Teachers, Successful Students", building on the World Development Report, 2018. The tool was designed for use at primary level, and in 2021 was expanded to early childhood education (ECE) through the launch of Teach ECE, which is suitable to use with 3-6 year olds and is fully aligned to Teach Primary to aid continuity of use. Teach Secondary was still in development at the time of writing.

Teach Primary tracks the amount of time that teachers and students spent on learning activities, and the quality of teaching practices as explored through three domains[18]:

- **Classroom culture**: Considers whether teachers create positive learning environments which are supportive and conducive to learning, and which set positive behavioural expectations.

- **Instruction**: Explores whether teachers adopt methods of instruction which deepen student understanding and encourage critical thought and analysis, by facilitating lessons, checking for student understanding, providing feedback, and encouraging critical analysis.

- **Socio-emotional Skills**: The teacher fosters socio-emotional skills such as autonomy, perseverance and social and collaborative skills.

Teach ECE similarly explores the same aspects of Classroom Culture and Socio-emotional Skills as in Teach Primary, but 'Instruction' becomes 'Guided Learning', to reflect the greater depth of interaction between teachers and younger children.

How does the tool work?

Teach comprises both high and low inference measures, with two main components of teacher effectiveness:

- **Time-on-Task observation**: Involves three 1-10 second 'snapshots' to record whether the teacher is providing a learning activity (scored as yes/no), and the extent that students remain on task (scored as low, medium or high).

- **Quality of teaching practices observation**: Involves a more detailed analysis of the three domains of classroom culture, instruction, and socio-emotional skills. Two 15-minute observations are conducted to explore nine elements and 28 behaviours. Each of the 28 behaviours are scored from low to high, based on a five point scale.
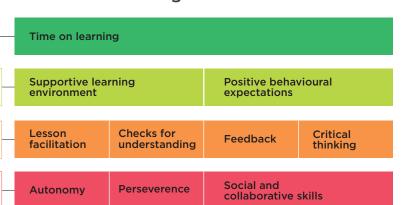
Figure 3 - Teach framework

Source: Molina *et al*, 2021

---

18   Molina *et al*, 2021

3 How to choose the right classroom observation tool?

Stallings, CLASS and Teach each offer a range of benefits, providing users with data and evidence to help understand what is happening in classrooms, and generate insights to inform policy, practice, and teacher professional development. However, not all tools are equally suited to different purposes, research aims or audience needs, and it is useful to consider the drawbacks of each tool when deciding which one to select. Selecting the most appropriate tool for your project or intervention will depend on a range of factors, including the type of information required, how findings will be used, and the time and resources

available to conduct the observations, analyse the data, and apply the findings.

This section draws on existing literature, and the experiences of individuals interviewed during the development of this guide, to set out some of the key advantages and limitations of each tool. We consider: ease of use; costs; the strength and reliability of the data produced by each tool; their adaptability to LMIC contexts; and the potential to explore global priority issues surrounding gender, disability and socio-emotional learning.

3.1 Stallings

When and why should I use it?

When rigorous and comparable 'snapshots' of what is happening in classrooms are required

• • •

When easily communicable data are needed to inform and influence policy and practice

• • •

When the priority is to conduct large scale research across a range of contexts in a cost effective way

• • •

What to consider?

Caution required when reviewing results for *individual* teachers

• • •

Cannot measure the *content* of what is being taught

• • •

Not sufficient by itself to explain learning *outcomes*

• • •

Does not provide insights into *gender, disability or socio-emotional learning*

• • •

What are the advantages of Stallings?

The main benefits of Stallings are derived primarily from its simplicity – it generates rigorous and comparable quantitative data that can easily be communicated to policy makers and decision makers, and can be delivered cost effectively on a large scale[19]. These points came out consistently in our literature review and during consultation calls[20].

• **Generates rigorous and comparable data**: Stallings can provide powerful insights where little is known about what is happening in classrooms. Results are comparable across schools, regions, or countries, and can be used to develop indicators that benchmark and track progress overtime[21]. The tool has been

extensively validated, and achieves high levels of inter-rater reliability between observers. This means there is high level of consistency between different raters' decisions or ratings, so the manner in which they rate different observations is likely to be similar.

• **Easy to analyse and communicate findings**: Due to the binary nature of Stallings' data, analysis is straightforward, and findings can be easily communicated to inform and influence educational reforms and priorities. One expert consulted in the development of this guide stressed that this data can act as an important wake up call for Ministries of Education,

19   World Bank, 2017; Bruns and Luque, 2015
20   Two experts consulted had significant experience using Stallings and analysing Stallings data
21   Bruns *et al*, 2016

highlighting where there are major gaps in the time spent actually teaching in class, and where students are disengaged or absent. Research has also highlighted the stark difference in results across schools and contexts, helping policy makers, school leaders or others involved in education delivery to identify where further support and resources are required.

- **Cost effective to deliver on a large scale**: The tool is well suited to large scale data collection, and achieves valid results without extensive (and expensive) training of observers, even when they have limited experience of education delivery[22]. It is both language and curriculum neutral, so the tool can be used in any school or context, and results remain comparable both within and across countries[23]. Stallings is generally conducted in person, but can feasibly be carried out remotely using video recordings. All these factors make it an attractive option for large-scale research in LMIC settings.

What should be considered before using the Stallings instrument?

Although Stallings data is highly regarded and can provide powerful insights into what is happening in classrooms, it is important to note its limitations. The Stallings user guide, developed by the World Bank, highlights some issues to consider before using this tool:

- **Understanding classroom processes rather than *individual* teacher performance**: The validity of the tool in measuring individual teacher performance is less clear, due to the number of variables that may impact a teacher's performance in any one class, and which may change from class to class, or day to day. The tool cannot account for the composition of students in individual classrooms, nor does it provide an in depth assessment of teaching practices. Therefore, it is not possible to know whether teachers are managing classrooms more effectively, or if students are higher ability and more motivated to participate and engage. It is therefore not recommended as a high stakes, individual performance assessment tool, and teachers should remain anonymous in analysis and reporting.

- **Limited insights about the *quality* of teaching content**: While the subject and curriculum neutral nature of Stallings means the tool can be used across different grades, subjects and contexts, it cannot measure the quality of teaching *content* itself. Two teachers may appear very similarly in Stallings metrics – having the same ratio of time-on-task and using learning materials to the same degree – however they may be delivering teaching of qualitatively different levels due to the nature of their interactions with students[24].

- **Focus on activites rather than learning outcomes**: Used in isolation, Stallings does not provide detailed information on teacher-student interactions. This means it is not always sufficient by itself to explain student learning outcomes. Although there is a high correlation between a high score on Stallings and improved learning outcomes, this is not always the case. As a counter-example, teachers in Afghanistan have been found to have low absence rates and relatively high time on learning activities – at 80% it is close to the target of 85% – but because of their very low content and pedagogical knowledge, students are not necessarily learning[25].

How adaptable is Stallings to low and middle-income settings?

Although Stallings was designed in and for the US context, it has been validated in several LMICs and is arguably well-suited for large-scale use across a range of contexts[26]. The tool is language and curriculum neutral, and measures aspects of teaching that are minimally affected by cultural factors. The straight forward nature of the Stallings tool is useful for informing education system policy and practice based on large scale and representative samples[27], and as mentioned above, ensures it can be delivered cost effectively on a large scale.

22   Abadzi 2007; Schuh Moore, DeStefano, and Adelman 2012.
23   World Bank, 2017
24   World Bank, 2017
25   Molina, *et al*, 2018
26   Bruns *et al*., 2016; World Bank, 2017; Hertz *et al*., 2019
27   Bruns *et al*., 2016; World Bank, 2017

Exploring teacher classroom practice in seven Latin American and Caribbean countries

The World Bank collaborated with seven Latin American and Caribbean (LAC) countries - Brazil, Colombia, Honduras, Jamaica, Mexico, and Peru and a pilot effort in the Dominican Republic - between 2009 and 2013 to collect data on classroom practices from over 15,000 classrooms using the Stallings instrument.

The research had three main objectives:

- Benchmark the current performance of LAC's teachers and identify key issues.
- Share emerging evidence on important reforms of teacher policy being implemented in LAC countries.
- Analyse the political 'room for manoeuvre' for further reform in LAC.

The findings from the study highlighted low levels of instructional time on average across all countries, with most lost time used on classroom management, or teachers being completely off task. Teachers often did not use school resources available to them, and students were regularly disengaged from teacher-led activities. Finally, the study found a significant amount of variation both across schools and within schools.

Findings from the study led to several partner countries taking action to address teacher effectiveness at the classroom level. This included mainstreaming periodic classroom observations using the Stallings instrument into regular school supervision processes, reforming teacher hiring processes, and revisiting in-service teacher development programmes. In some cases, videotapes of high performing teachers are being shared with teachers as examples of 'good' teaching practices, or using video to allow teachers to see and analyse their own teaching practice.

What insights can Stallings provide on issues of gender, disability and socio-emotional learning?

Stallings does not (and was not designed to) provide insights on gender, disability and socio-emotional learning. The Stallings protocol calls for the observer to count the number of students in each class at the beginning and end of the lesson, which could provide an opportunity for gender disaggregated analysis. However, this does not allow for any analysis of gendered interactions between teachers and pupils, and it is not recommended as the sole measure of gender parity and attendance. Similarly, the tool is not designed to provide information on the individualised support and resources available for children with disabilities. Finally, the simple nature of the tool is not suitable for exploring more complex and nuanced interactions between teachers and students, which can support the development of socio-emotional learning and life skills.

> The Stallings protocol calls for the observer to count the number of students in each class at the beginning and end of the lesson, which could provide an opportunity for gender disaggregated analysis.

When and why should I use it?

When an in-depth analysis of teacher-student interactions and teaching processes is required

• • •

To provide teachers with individualised feedback to improve teaching practice

• • •

When interested in identifying teaching practices that have a proven link to students' academic and socio-emotional outcomes

• • •

What to consider?

Not an open-source tool therefore more expensive and less accessible

• • •

Requires administration by certified observers who undergo paid training requiring sufficient time, resources, and budget

• • •

The ability of teachers to fully comprehend and act upon feedback may be limited in some contexts

• • •

What are the advantages of CLASS?

CLASS focuses on more qualitative and in-depth measures of teacher-student interactions, and while the instrument is less straightforward to implement and analyse, it provides valuable information about teaching practice.

- **Analyses more complex teaching processes and interactions**: CLASS looks at *how* teaching processes drive learning, by focusing on the interactions between students and teachers that effect student wellbeing and build the skills needed for future success[28]. In this way, it provides greater depth of information around teacher quality, and may have greater potential to explain the link between teaching practices and learning outcomes – although it should be noted that 'learning' itself is not something that can be directly observed[29], and so evidence of this would need to be supplemented with more robust measures of learning.

- **Measures teaching behaviours which have proven links with student achievement**: Research conducted on the validity and reliability of the tool have found that CLASS measures are positively associated with student achievement, with students having higher levels of learning when taught by teachers with high observation scores[30]. This is supported by multiple randomised controlled trials which linked teacher scores with students' socio-emotional and academic outcomes[31]

- **Provides individualised feedback to improve teaching practice**: It is a particularly useful tool for identifying and providing teachers with individualised and tangible feedback to improve their teaching instruction, focused on teaching practices that can improve learning and classroom experiences[32]. For this reason CLASS is widely used as a professional development tool in the US[33].

- **Wide coverage of age groups**: CLASS has been adapted for use with a wide range of age groups, from infant through to secondary level. The tool was initially developed as part of studies looking at early childhood care, and now covers six protocols covering birth through to secondary level[34]. Each tool is tailored to explore the unique needs of each age group, and provides definitions of effective teaching across the different education levels[35].

• •

28 Teachstone, 2018
29 Ofsted, 2018
30 Kane and Staiger, 2012
31 Thames and Hayashi, 2018
32 Bruns *et al*, 2016
33 Teachstone, n.d.c
34 Teachstone, n.d.b
35 A second edition of the CLASS measurement suite designed for early childhood care settings up to early grades (pre-K to third grade) is planned for launch in July 2022 at the time of writing (Teachstone, 2022)

What should be considered before using CLASS?

The complexity of the CLASS tool brings some drawbacks for certain users, primarily the cost and time associated with delivering the tool in lower resource settings.

- **Costs**: CLASS is a trademarked tool and is not open source, therefore there are greater costs associated with accessing and delivering the observations, which may make it less attractive to users with limited resources. Observers are required to undergo paid training and certification provided by Teachstone or authorised training partners, available in Australia, Canada, Chile, Colombia, Norway, Peru, Puerto Rico, and the USA, as well as online. Annual recertification is recommended to ensure observers continue to use the measure fairly and accurately.

- **Requires experienced and skilled observers**: To ensure observers are able to make consistent qualitative judgements when administering CLASS, it is important to identify and train individuals with the necessary skills and ability to complete the CLASS training. This level of experience is not always easy to identify in certain contexts, and may limit its usage in larger scale research[36]. One solution is to record videos of lessons, and have a smaller number of observers code the recordings – however some of the experts consulted in this research noted that not all student interactions are easily captured through recorded lessons.

- **Limited ability of teachers to respond to feedback in certain contexts**: A final consideration is whether the teachers being observed are in a position to act upon the feedback provided. There may be limited professional development tools available to help teachers analyse the results and identify strategies for improvements, or they may be grappling with more fundamental challenges with time and classroom management. These are often systemic issues which are outside of their direct control, and can limit their opportunities to focus on complex and nuanced student interactions[37]. For this reason, it would be important to consider whether the increased costs and resources to deliver CLASS are justified.

How adaptable is CLASS to low and middle-income settings?

CLASS has been validated in various contexts, including the US, China, Latin America, and Israel. It is designed to investigate 'universal' teaching practices, those that are relevant across cultures and contexts. Consequently, there is flexibility within the tool when it comes to examples of behavior that observers should be aware of. This is important to consider when using the tool in cross-country settings, as practices might be expressed differently in different contexts - for instance, play can be manifested differently across cultures, and norms around 'good behaviour' in one context may look quite different in another.

As highlighted above, the costs and resources involved in delivering CLASS is likely to be a key drawback for those looking to deliver the tool on a large scale, and the complexity may not be appropriate to provide individualised feedback to teachers in all contexts. These factors should be considered before adopting this tool in LMIC contexts.

36 Bruns *et al*, 2016
37 Bruns *et al*, 2016

The Quality of Early Childhood Care Centers (Centros Infantiles del Buen Vivir) in Ecuador

The InterAmerican Development Bank collected information from a sample of 404 public child care centres in Ecuador in 2012. The objective of the study was to characterise different quality aspects of the services provided by these centres to children under 3 years of age. CLASS was incorporated as one of the key data collection instruments.

A pair of researchers were assigned to each child care centre. Each pair included an experienced interviewer, who led the interviews and filming, and a researcher with post-secondary education in the field of child psychology or early childhood education, who administered the observational instruments. Each pair of researchers was assigned a field supervisor who was trained on all of the instruments, and whose specific task during the data collection process, in addition to logistical organisation, was to administer a questionnaire on infrastructure variables. All of the instruments were administered in the same order and at the same time of day to ensure comparability across centres.

The study was able to analyse various aspects of both structural and process quality, and found that process elements in particular require further improvement. The CLASS tool highlighted low levels of quality scores across nearly all the centres studied. The quality challenges faced by these centres are consistent with those of similar services in different countries throughout the region, such as Peru, Colombia, and Brazil. The study also found that while efforts have been made in recent years to improve service quality in Ecuador, continuous and long-term interventions will be necessary to achieve significant improvements in the quality indicators studied.

What insights can CLASS provide on issues of gender, disability and socio-emotional learning?

As highlighted above, one of the key strengths and design functions of CLASS is that it explores how teaching instruction can support students to develop socio-emotional skills which set them up for future success. This includes exploring teaching practices that can influence student autonomy, leadership, respect, self-worth, critical thinking and problem solving abilities.

It should be noted that the CLASS tool is not intended to *measure* SEL skills, but acts as a *predictor* of such skills in relation to teaching practices. During the consultation process for this report, one participant urged caution in blanket assumptions that CLASS can predict SEL, and noted the nuances of how different domains of emotional support, classroom observation, and instructional support are differently linked to different outcomes. A study looking at CLASS observations and early adolescent emotional and behavioural engagement explored how different dimensions of CLASS were differently associated with student engagement – i.e. certain teaching practices are more strongly associated with student engagement than others[38]. This nuanced understanding of how different dimensions affect engagement may enable users of CLASS to better support teachers to improve specific practices that are more strongly associated with specific socio-emotional skillsets.

Our review of existing literature found limited examples of research which has explored the reliability or validity of CLASS to predict student SEL outcomes in low and middle income contexts specifically. However, previous studies have

38   McKellar *et al*, 2019

indicated that it is applicable in diverse settings within the US context, such as classrooms with high proportions of dual language learners[39]. It may be necessary to conduct further research to explore the applicability of CLASS as a predictor of SEL skills in non Western contexts.

At present, the CLASS instrument is not designed to explicitly explore disability or gender. Currently, CLASS has been used in inclusive settings, which have a mix of children with and without disabilities, but observers are not trained to specifically identify gender or disability sensitivity of teaching methods.

"

One of the key strengths and design functions of CLASS is that it explores how teaching instruction can support students to develop socio-emotional skills which set them up for future success.

39   Downer *et al*, 2011

3.3 Teach

When and why should I use it?

When users would benefit from a holistic measure of what is happening in the classroom in terms of both time-on-task and teacher behaviours and practices

• • •

When users are particularly interested in focusing on inclusive teaching practices to increase opportunities for all children to learn

• • •

The open-source tool is easily accessible to anyone wishing to explore teaching quality in the classroom.

• • •

What to consider?

It is not intended for use in high-stakes decision-making or to evaluate individual teachers

• • •

Appropriate time and resources should be allocated to sufficiently train observers, adapt the tool to the local context through the use of local videos, and test for inter-rater reliability

• • •

Findings from validation studies are promising, but evidence is still emerging

• • •

What are the main advantages of Teach?

Teach incorporates both the in-depth benefits of high inference tools, along with low inference measures of time on task, offering a 'best of both worlds' option.

- **Holistic measure of what is happening in the classroom**: A key selling point of Teach is it's ability to shed light both on quantitative measures of the time teachers spend on task, in addition to in-depth measures of teaching quality which can contribute to both academic and socio-emotional learning[40]. This may be an attractive option to those who wish to collect easily quantifiable measures of what is happening in classrooms, without compromising on exploring more complex individual teacher behaviours or practices linked to academic and social learning outcomes[41].

- **Designed to explore inclusive teaching practices**: Teach has been specifically designed to explore inclusive teaching practices by measuring whether teachers exhibit bias or challenge gender or disability stereotypes. It has also recently been revised to strengthen this aspect of the tool. In 2021, the instrument was revised to include an additional 'behaviour' to measure disability bias, revisions were made to existing behaviors to better capture inclusive teaching practices, and supporting materials were updated to provide stronger inclusion

focused examples. See more on the inclusive elements of the tool below[42].

- **Open source tool, manual, and resources available in multiple languages**: Teach was designed to be accessible to anyone wishing to generate insights into teaching practices, and therefore the tool can be accessed for free, along with a detailed instruction manual and wide range of instructional materials and resources such as videos, templates, and guidance notes on how to plan for observations, adapt the tool, and analyse and report the results. It is available in English, Arabic, French, Portuguese, Russian, Spanish and Swahili.

What should be considered before using Teach?

Teach is a relatively new tool, and therefore research exploring its relevance and effectiveness as a measure of teaching practices in LMICs is still emerging. Some urge caution in how the tool is used, highlighting some considerations, such as:

- **Teach is not intended to be used to evaluate individual teachers**: The World Bank has stressed that the purpose of the tool is *not* to evaluate individual teachers per se, and certainly not for high stakes assessments. Data should inform policymakers about the quality of interactions between teachers and students, and to track the effectiveness of teacher reforms[43]. It might also be useful to consider that the lack

40 World Bank, 2021a
41 Carter *et al*, 2020a
42 See also the World Bank (2022) Teach website for more information: https://thedocs.worldbank.org/en/doc/
    b1edabf99b0a079db4708327b12676e9-0200022022/related/2022-TeachPrimary-UpdatestoToolSecondEdition.pdf
43 Edwards, 2019

of immediate feedback provided by observers (who may not be trained teachers) limits the opportunities for teachers to engage or discuss their own reflections during the observation process[44]. While Teach can also be used as a professional development tool to identify the strengths and weaknesses of individual teachers, this should be done in conjunction with materials from the Coach programme, focused on helping countries improve their in-service teacher professional development programs and systems. The Coach programme includes a range of resources and tools that build on Teach and help countries develop systems to support teachers in improving specific teaching practices[45].

- **Users should allow sufficient time and resources to pilot and adapt the tool**: While the Teach tool is open source and therefore presents a cost effective option for those using classroom observations, it is essential that sufficient time and resources are allocated to reviewing the contextual relevance of the tool with in-country stakeholders, that observers are fully trained, and that the tool is piloted and tested for inter-rater reliability to ensure it is used and interpreted effectively. The scoring criteria, for example, may be open to interpretation in some instances, therefore users should test for consistent understanding and scoring across observers to ensure findings are valid and reliable[46]. As highlighted by some commentators, there may be a risk that users with limited budget and tight timelines may assume the tool is reliable in all contexts, and be tempted to skip the adaptation, piloting and inter-rater reliability assessment stages[47].

- **Findings from validation studies are promising, but evidence is still emerging**: It is important for those using the Teach tool to be conscious that it is still a relatively new instrument, having only been launched in 2019. Therefore, while there is promising research into the reliability and validity of the tool, evidence of its relevance across different countries and contexts is still emerging, and continuing to generate further insights and evidence on the links between Teach scores and student learning will be an important future priority.

How adaptable is Teach to low and middle-income settings?

Unlike tools designed for the US and later adapted and revalidated for use elsewhere, Teach was designed specifically for use in LMIC contexts. Prior to roll out, Teach was piloted in more than 1,000 classrooms in Mozambique, Pakistan, the Philippines and Uruguay, and global video footage tested in 11 LMICs. Analysis of this pilot data found correlations between teachers who score highly in Teach and students with high learning outcomes. The thorough testing of reliability and validity has demonstrated that it can achieve high levels of inter-rater reliability across different contexts, and Teach has now been launched in 36 countries[48].

In addition to Teach, the World Bank has developed complementary resources as part of the Coach initiative, which provides support to countries to implement and improve a range of in-service teacher professional development modalities to enable users of classroom observations to act upon the data generated through Teach[49].

It is important to stress that while the tool is designed to be adapted for specific contexts, users must build in sufficient time to test and pilot adapted resources effectively to ensure the tool remains reliable. Inter-rater reliability assessments will ensure that observers have a consistent understanding of what is being measured, which can change from context to context, and as such reliability testing remains an important step[50].

Although Teach is open source, users should factor in costs associated with observers' time during training and research. Costs will depend on the approach taken to data collection, such as whether it is conducted in person or via video recordings. Translation and language adaptation are another factor to consider - interviewees remarked that while the materials were generally accessible to wider audience, some adaptation was needed to further simplify or adapt the language, and research conducted by the Research for Equitable Access and Learning (REAL) Centre has stressed the time required to adapt the tool and adequately train observers was extensive, but essential[51].

· ·
44 Pesambili *et al*, 2022
45 World Bank, 2021b
46 RTI, 2016
47 Jukes, 2019
48 World Bank, 2021a
49 See also World Bank 2022, available at: https://thedocs.worldbank.org/en/doc/786b435688a243c8b7a6f3014055edfe-0140062021/related/Teach-in-Action-Final-Ana-Teresa-Del-Toro.pdf
50 RTI, 2016
51 Carter *et al*, 2020b

Using the Teach tool in Rwandan secondary schools[52]

Teach was recently used in Rwanda to observe video recordings of lessons in 103 secondary schools, and track changes in teaching quality in a teacher professional development programme. In consultation with Rwandan teachers, the tool was adapted in advance of data collection to ensure it was relevant to the country context. Some examples of adaptations include:

- **Cultural adaptations**: Including specific culturally relevant behaviours, e.g. making a gesture of a flower to show positive feedback to students, a common practice in Rwanda and used as an example of positive language.

- **Educational**: Nuancing behaviours to reflect a secondary context, e.g. changing all references of students using pencils in the classroom to pens, as students no longer use pencils in the secondary context.

- **Programme**: Focusing examples on STEM subjects at the secondary level, e.g. examples refer to more advanced work in mathematics subjects such as statistics and probability, rather than basic addition and subtraction.

The study provided practical reflections and recommendations for those looking to use of the Teach tool in international settings:

- Adapting Teach helps capture context-relevant aspects of teaching that play an important role in assessing teaching quality. Applied by trained researchers, the adapted tool provides reliable insights on teaching quality specific to the Rwandan context.

- The time needed to train researchers to use and code the tool, particularly when working with an adapted version for the first time, is significant.

- There are pros and cons of using video for coding observations. For example, student and teacher dialogue can be difficult to discern, however, video allows for cross-checks by other researchers and removes some of the burden from the researcher to capture everying that is happening in real time.

- It is recommended that users build in feedback loops so that the findings are shared with teachers, implementing partners of interventions, and policy-makers in a practical and constructive way to improve teaching and learning outcomes.

Further rounds of data collection are planned, to explore whether students taught by teachers with higher scores on the tool perform more highly and also to track changes in classroom practices.

52   Carter *et al.*, 2020b

What insights can Teach provide on issues of gender, disability and socio-emotional learning?

Teach is one of the few classroom observation tools that has been specifically designed to analyse issues associated with both **gender** and **disabilities**, as well as other dimensions of inclusion such as language of instruction. Recent adaptations to the tool and associated resources have further strengthened this element. For instance, the tool manual includes an optional checklist to assess the accessibility of the physical environment[53]. They allow users to investigate: forms of representation in the classroom; the prevalence of gender or ability bias from teachers or in educational resources; how teachers adjust for different learning needs; and the availability of specialised support or learning aides[54]. The training guide provides advice and recommended activities to ensure there is shared understanding across observation teams on what constitutes gender bias and gender stereotypes[55].

Teach has also been specifically designed to cover issues of **socio-emotional learning**. These are addressed under the three elements of:

- Autonomy;

- Perseverance; and

- Social and Collaborative Skills.

'**Autonomy**' asks that observers code against whether teachers provide students with opportunities to make choices, and whether students volunteer to participate. '**Perseverance**' explores the extent that teachers recognise student efforts, have positive attitudes towards their challenges, and encourage goal setting. Finally, '**Social and Collaborative Skills**' checks whether whether teachers promote peer interaction and interpersonal skills, and whether students actively collaborate with one another[56]. Teach Secondary has a greater focus on socio-emotional skills, due to their positive associations with cognitive development and importance for employability[57].

> ❝
>
> The Teach training guide provides advice and recommended activities to ensure there is shared understanding across observation teams on what constitutes gender bias and gender stereotypes.

53   See p 44 of the Teach Observer Manual, available here: **https://documents1.worldbank.org/curated/en/872291641201520569/pdf/Teach-Primary-Observer-Manual.pdf**
54   World Bank, 2019; World Bank, 2022; **https://www.youtube.com/watch?v=-2WaFdhhEQs&t=1256s**;
55   See also World Bank, 2022, available at: **https://thedocs.worldbank.org/en/doc/b1edabf99b0a079db4708327b12676e9-0200022022/related/2022-TeachPrimary-UpdatestoToolSecondEdition.pdf**
56   World Bank, 2022
57   Carter *et al.*, 2020a

4 Reflecting on the impact of COVID-19

When COVID-19 was declared a pandemic in March 2020, many countries closed their schools. While education systems in higher-income countries delivered online teaching, classroom observations, including those of video-taped remote lessons, were suspended for the duration of the school closures. In lower-income countries, teaching itself was suspended as there were few to no resources for remote teaching such as lesson worksheets, and the ICT infrastructure did not have the capacity for distance learning[58].

During the lockdowns and as countries started to reopen the schools, it was recognised that schools and teachers need help to address any learning loss. Hence, the emphasis that most stakeholders were putting on observing learning was not on teacher evaluation, but on supporting teachers to address the multiple challenges raised by COVID-19, including those to students' and teachers' wellbeing. As a result, the use of the tools presented in this guide was somewhat
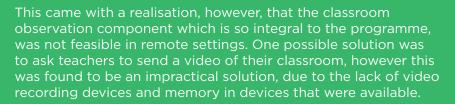
limited during the pandemic. Many organisations have turned their attention to providing ongoing support for the educators during this emergency situation. In some cases, organisations have adapted from using classroom observation to monitoring the learning in alternative ways, given that pandemic and schools closures meant that previously-used classroom observation approaches were not possible.

With the schools reopening there is, however, an increased interest in understanding how classroom learning can continue after the disruptions and how systems can build back better. In this context classroom observations tools can deliver important insights into understanding how teaching and learning looks after Covid-19. This can be particularly useful in situations where a comparison is possible between observations pre- and post-school closures.

Monitoring classrooms during COVID-19 – an example from Kenya

Dignitas, an educational organisation working directly with schools in Kenya and South Sudan, typically employs modified classroom observation tools created to fit its specific purposes and goals in order to monitor and support school leaders and teachers. Since COVID-19 resulted in national school closures in Kenya in mid-March 2020, the programme adapted it's activities to coaching virtually and remotely. First, it identified the competencies needed to support teams to deliver education remotely and ran a virtual training and coaching programme for school leaders for six months. In the second phase of the COVID-19 response, Dignitas identified competencies for reopening and recovery, with a focus on learner and teacher wellbeing.

This came with a realisation, however, that the classroom observation component which is so integral to the programme, was not feasible in remote settings. One possible solution was to ask teachers to send a video of their classroom, however this was found to be an impractical solution, due to the lack of video recording devices and memory in devices that were available.

In the absence of classroom observations, and to support the identification of teacher competencies, Dignitas partnered with the University of York to develop Scenario Based Learning (SBL) approach, which comprises situational judgement tests to gauge school leaders' and teachers' competencies, and their responses to different scenarios. Whilst SBL is not a complete replacement for classroom observations, it does provide coaching data and insights into the shift in competencies in various areas of teaching practice, including learner engagement and differentiated support.

For more information on Dignitas, visit: **dignitasproject.org**

Our comparison reflects the relative strengths and limitations of the three tools, which may influence which tool is appropriate for any given context. Previous research on teacher effectiveness comparing the tools found there was a consistency across each one in terms of the measurement of time spent on tasks in classrooms, but less consistency in measures of other (even thematically-similar) aspects of teaching. This suggests that the choice of tool might also shed light on different aspects of teaching and its effectiveness[59].

Table 2 offers a comparison of the tools at the glance. Taking all the information together, a few main aspects emerge:

Stallings

- Captures basic information
- Open source
- Four languages
- Limited options to investigate interactions regarding gender, disability, and socio-emotional learning

Stallings is excellent for capturing basic information teaching practices, such as time-on-task and student engagement, it is open source, and available in four languages. It does explicitly address the issues of gender, disability, and socio-emotional learning.

CLASS

- Captures richer data
- Requires training at a cost
- Three languages
- Covers all age ranges

CLASS captures much richer data, but is therefore more complicated to use and requires training and certification. This comes at a cost, but also includes professional support throughout the research phase. It is the only tool reviewed that covers all age ranges from birth through to secondary level, and it is available in three languages.

Teach

- Captures low- and high-inference data
- Open source
- Multiple languages
- Covers early childhood and primary, secondary coming soon

Teach was specifically designed for use in low- and middle-income countries and collects both low- and high-inference data, which addresses some of the shortcomings of Stallings while remaining accessible for users. It currently covers early childhood and primary, and a tool for secondary will be launched in the near future. It is open source and available in multiple languages.

59   Filmer *et al*, 2020

Table 2 - Comparison of Stallings, CLASS, and Teach at a glance

	Stallings	CLASS	Teach
Education levels	Primary and secondary	Birth through to secondary	Primary and Early Childhood (secondary coming soon)
Type of data collected	Low-inference	High-inference	High-and low inference measures
Contexts	Designed in and for US but widely used in LMICs	Designed in and for US but widely used in LMICs	Designed for use in LMICs
Availability	Open source	Requires training + certification	Open source
Training recommended[60]	» 2.5 days	≥ 2 days	» 4 days
Resource consideration	Staff and transport costs	Staff and transport costs Training costs, training includes support during research phase	Staff and transport costs. Possible adaptation costs
Languages	English, French, Portuguese, Spanish	English, French, Spanish	English, Arabic, French, Spanish, Portuguese, Swahili, Russian, with more languages available for certain editions of the tool
Gender	Not covered	Not covered	Gendered teaching explored
Disability	Not covered	Limited	Strengthened focus on disability in second edition
Socio-emotional learning	Not covered	Explores behaviours linked to SEL outcomes	Explores behaviours linked to SEL outcomes

60 The length of the training can have impact on the reliability with which the trained participants can reliably code the tool in question. Some evidence suggests that longer training can be helpful in ensuring a higher reliability in coding, both for the individual (intra-reliability) as well as between them (inter-reliability). This might be worth considering, especially in larger-scale administration of the tool and included as an area of investigation to assess the training.

As noted, all three tools are constantly evolving. There are new versions of CLASS and Teach in development at the time of writing.
It is therefore important to check the relevant resources for updates before making a decision on a particular tool.(see Section 6 for a list of resources).

Some of them are part of larger initiatives aiming to build successful teaching and learning. For instance, the World Bank Coach tool aims to help countries improve teacher professional development, and helps countries to leverage insights from Teach data 'to tailor the support teachers receive to improve their teaching'. This includes open-access tools and resources for 'policymakers, researchers, school and system leaders, coaches and teachers'[61].

What are the limitations of classroom observations?

While classroom observation tools can provide valuable insights for education policymakers and practitioners, it is important to also recognise the limitations and critiques of these standardised approaches when used across multiple contexts. These limitations are present across all three tools reviewed, including:

- **Limited view of teaching practices**: While the tools discussed observe and score a range of practices, for practical reasons these tend to be limited to a set of pre-defined indices. These reduce the complexity of teacher work and attempt to quantify the complex social dynamics that play out in teaching and learning environments[62]. The benefit of this is that scores can be tested for reliability and validity across observers, but it can risk losing the context in which behaviours play out and may not provide a full and nuanced picture.

- **Tools reflect specific points in time**: Many classroom observations, including those explored in this paper, generate a brief snapshot of what is happening in the classroom during a short timeframe. Results are therefore dependant on the amount of time the
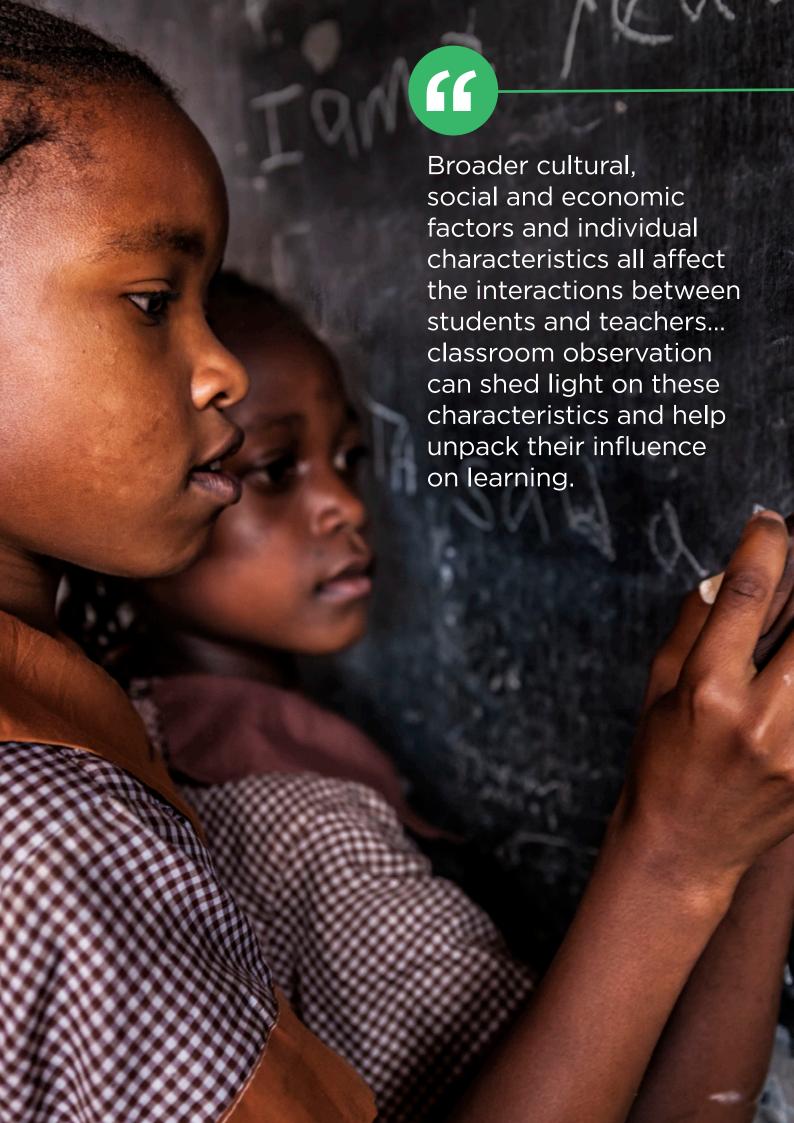
observation takes place and the contextual factors that affect a teachers performance in that moment. It is therefore possible that teachers are not able to showcase particular skills during the timed observation. Given the range of potential factors that can impact on a teachers' practices and behaviours at any given time, particularly in LMIC contexts, it is generally not recommended that classroom observation findings are used for high stakes assessments of teachers.

- **Social desirability**: If teachers and students are aware of the observations and if these are treated as 'high-stakes' (e.g. if pay increase is associated with these) there is a potential for teachers to change their behaviour in line with what they think might be expected of them, rather than being a true reflection of average teaching behaviours. This should be taken into account when analysing and interpreting findings.

- **Concerns about globalised and Westernised 'one size fits all' approach**: Many classroom observation tools, particularly those which provide global, comprehensive and standardised protocols for use in multiple contexts (such as those explored in this paper) are critised for applying Westernised ideas of concepts and understandings of 'good' and 'bad' teaching practices. As highlighted in previous sections, it is important for users to consider how they are adapting and contextualising tools, including ensuring the relevant stakeholders (including teachers themselves) are engaged and informed when adapting tools for the local context. Among the reviewed tools, the reviews of Teach have notably included evidence from different world regions, including low and middle-income contexts.[63]

Despite these limitations, classroom observations remain a useful and popular tool for exploring teaching quality, depending on the information needs of the user. When combined with other research methods, they have the potential to generate insights to identify and improve teaching practices which influence student learning.

61   World Bank: https://www.worldbank.org/en/topic/teachers/brief/coach-helping-countries-accelerate-learning-by-improving-in-service-teacher-professional-development [Accessed 22 March 2022]
62   Pesambili *et al*, 2022
63   Examples can be found in the review for the Teach early childhood education tool: https://thedocs.worldbank.org/en/doc/75bdb5f2c03f19f0642db1c941193f8d-0140042021/related/Teach-ECE-Lit-Review-5-27-21-clean-Template-3.pdf

> Broader cultural, social and economic factors and individual characteristics all affect the interactions between students and teachers... classroom observation can shed light on these characteristics and help unpack their influence on learning.

Abadzi, H. (2007). *Absenteeism and Beyond: Instructional Time Loss and Consequences*. World Bank Policy Research Working Paper No. 4376.

Bruns, B., & Luque, J. (2015). *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean.* Washington D.C.: World Bank.

Bruns, B., De Gregorio, S., & Taut, S. (2016). *Measures of Effective Teaching in Developing Countries.* Oxford: RISE Programme.

Carter, E., Molina, E, Pushparatnam, A. & Rose, P. (2020a). Measuring Teachers' Encouragement of Socioemotional Skills in the Secondary Classroom. In: Smart, A., and Sinclair, M. (Eds.). (2020). *NISSEM Global Briefs: Educating for the Social, the Emotional and the Sustainable: Pedagogy, Practice and Materials*. NISSEM.

Carter, E., Rose, P., Ani-Asamoah, B., Leonard, P., & Colenbrander, K. (2020b). *Video Recordings of Classroom Observations: Using the Teach Tool in Rwandan Secondary Schools*. Research and Policy Paper No. 20/9, REAL Centre, University of Cambridge.

Downer, J.T., López, M.L., Grimm, K.J., Hamagami, A., Pianta, R.C., Howes, C. (2011). *Observations of Teacher–Child Interactions in Classrooms Serving Latinos and Dual Language Learners: Applicability of the Classroom Assessment Scoring System in Diverse Settings*, Early Childhood Research Quarterly, Volume 27, Issue 1, Pages 21-32, ISSN 0885-2006, https://doi.org/10.1016/j.ecresq.2011.07.005.

Edwards, S. (2019*). 'Teach' Tool Not for Evaluating Teachers, says World Bank Education Boss*. Devex. Accessed June 23 2022: https://www.devex.com/news/teach-tool-not-for-evaluating-teachers-says-world-bank-education-boss-94362

Filmer, D., Molina, E. Wane, W. (2020). *Identifying Effective Teachers: Lessons from Four Classroom Observation Tools.* RISE Working Paper Series. No. 9365. World Bank, Washington, DC. © World Bank. Accessed June 23 2022: https://openknowledge.worldbank.org/handle/10986/34382

Hertz, A., Kochetkova, E., Pflepsen, A. (2019). *Classroom Observation Toolkit for Early Grade Reading Improvement*. USAID. Accessed June 23 2022: https://www.edu-links.org/sites/default/files/media/file/Classroom%20Obs%20Toolkit%20_Nov%202019.pdf

Jukes, M. (2019). *The Teach tool – the Start of a Global Collaborative to Improve Classroom Observations?,* RTI website, Accessed June 23 2022: https://shared.rti.org/content/teach-tool-%E2%80%93-start-global-collaborative-improve-classroom-observations

Kane, T.J., and Staiger, D.O. (2012). *Gathering Feedback from Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains*. Bill & Melinda Gates Foundation.

McKellar, S.E., Cortina, K.S,. Ryan, A.M. (2019) *Teaching Practices and Student Engagement in Early Adolescence: A Longitudinal Study Using the Classroom Assessment Scoring System*, Teaching and Teacher Education, Volume 89, 2020, 102936, ISSN 0742-051X, https://doi.org/10.1016/j.tate.2019.102936

MET Project. (2010). *The CLASS Protocol for Classroom Observations.* Seattle: Bill and Melinda Gates Foundation.

Molina, E., Trako, I., Matin, A. H., Masood, E., & Viollaz, M. (2018). *The Learning Crisis in Afghanistan: Results from the Afghanistan SABER Service Delivery Survey, 2017.* Washington D.C.: World Bank.

Molina, E,. Pushparatnam, A., Melo, C., Wilichowsk, T., Del Toro Mijares, A.T., Ding, E., Aloys, J., Carter, E., Singal, N. (2021). *Teach Primary Observer Manual (Second Edition)*. Washington, DC: The World Bank. License: Creative Commons Attribution CC BY 4.0 IGO.

Ofsted. (2018). *Six Models of Lesson Observation: An International Perspective.* Accessed June 23 2022: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/708815/Six_models_of_lesson_observation.pdf

Pesambili, J.C., Sayed, Y., Stambach, A. (2022). *The World Bank's Construction of Teachers and their Work: A Critical Analysis*, International Journal of Educational Development, Volume 92, 102609, ISSN 0738-0593, https://doi.org/10.1016/j.ijedudev.2022.102609.

RTI International. (2016). *Measures of Quality Through Classroom Observation for the Sustainable Development Goals: Lessons from Low-and-Middle-Income Countries.* Paris: UNESCO Global Education Monitoring Report.

Sanchez, H. (2021). *How Poverty and Stress Influence Students' Behavior.* Accessed 23 June 2022: https://www.ascd.org/el/articles/how-poverty-and-stress-influence-students-behavior

Schuh Moore, A-M., DeStefano, J., Adelman, E. (2012). *Opportunity to Learn: A High Impact Strategy for Improving Educational Outcomes in Developing Countries.* Washington, DC: EQUIP2, AED, and USAID.

Stallings, J., and S. Knight. (2003). *Using the Stallings Observation System to Investigate Time on Task in Four Countries.* Unpublished paper for the International Time on Task (ITOT) Project, World Bank, Washington, DC.

Teachstone. (n.d.a). *International*. Accessed June 23 2022: https://teachstone.com/international/

Teachstone (n.d.b). *CLASS Age Levels*. Accessed June 23 2022 https://info.teachstone.com/hubfs/CLASS%20Age%20Levels.pdf?hsLang=es

Teachstone (n.d.c). *The Classroom Assessment Scoring System® (CLASS®): Promoting Quality in Early Care and Education Settings across the Country.* Accessed June 23 2022: https://cdn2.hubspot.net/hubfs/336169/Teachstone%20CLASS%20System%20final.pdf

Teachstone. (2018). *Why CLASS? Exploring the Promise of the Classroom Assessment Scoring System® (CLASS).* Teachstone Training, LLC. Accessed June 23 2022: https://info.teachstone.com/hubfs/WhyClass_Ebook%20(1).pdf

Teachstone. (2022). CLASS® 2nd Edition: *A More Equitable, Accessible, and Impactful Future.* Accessed June 23 2022: https://teachstone.com/class-2nd-edition/

Thames, V., and Hayashi, R. (2018). *The Classroom Assessment Scoring System as a Measure of Effective Teaching to Address the Learning Crisis.* ADB Briefs (No. 92). Accessed June 23 2022: https://www.adb.org/sites/default/files/publication/432296/adb-brief-92-classroom-assessment-scoring-system.pdf

UNESCO. (2020). *Global Education Monitoring Report – Gender Report: A New Generation: 25 Years of Efforts for Gender Equality in Education*. Paris, UNESCO.

UNESCO, UNICEF, the World Bank and OECD. (2021). *What's Next? Lessons on Education Recovery: Findings from a Survey of Ministries of Education amid the COVID-19 Pandemic.* Paris, New York, Washington D.C.: UNESCO, UNICEF, World Bank

World Bank. (2017). *User Guide. Conducting Classroom Observations: Stallings "Classroom Snapshot" Observation System for an Electronic Tablet.* Washington D.C.: World Bank Group. Accessed June 23 2022: https://openknowledge.worldbank.org/bitstream/handle/10986/28339/119754-WP-PUBLIC-WB-Manual-ENG-WEB.pdf?sequence=1&isAllowed=y

World Bank. (2021a). *Teach Primary: Brief (Second Edition)*. Washington D.C.: World Bank Group. Accessed June 23 2022: https://thedocs.worldbank.org/en/doc/bd7b15d459fe78fab0201297b-9f4a45d-0200022022/related/Teach-Implementation-Brief-english-04012022.pdf

World Bank. (2021b). *Teach in Action: Three Case Studies of Teach Implementation*. Washington D.C.: World Bank Group. Accessed June 23 2022: https://thedocs.worldbank.org/en/doc/786b435688a243c8b7a6f3014055ed-fe-0140062021/related/Teach-in-Action-Final-Ana-Teresa-Del-Toro.pdf

World Bank. (2022). *Training Manual: Teach Primary*. Washington D.C.: World Bank Group. Accessed June 23 2022: https://thedocs.worldbank.org/en/doc/b1edabf99b0a079db-4708327b12676e9-0200022022/related/2022-TeachPrimary-TrainingManual-Public.pdf

Stallings

All the relevant Stallings resources can be found at https://www.worldbank.org/en/programs/sief-trust-fund/brief/stallings-program-download

CLASS

All the relevant CLASS resources can be found at https://teachstone.com/

TEACH

All the relevant Teach resources can be found at https://www.worldbank.org/en/topic/education/brief/teach-helping-countries-track-and-improve-teaching-quality
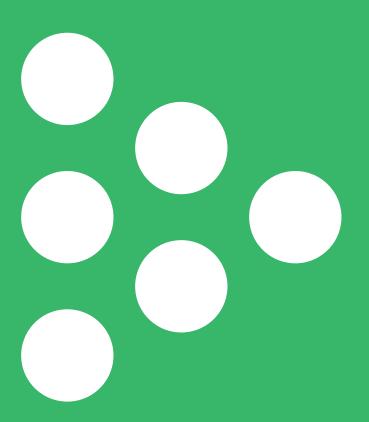
About NFER

NFER is a leading independent provider of education research, evaluation and assessment solutions worldwide. Our mission is to build better futures globally, and to support improvements within education systems.

We achieve this with our expert evaluations, research and assessments across all aspects of teaching and learning. This is used to inform policymakers and other education stakeholders worldwide, and also support the strengthening of classroom practice for all learners.
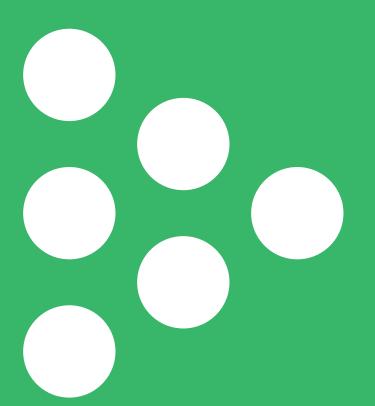
We work with a wide range of government departments and agencies, NGOs, donors, and civil society and private organisations. Our growing international portfolio of work has recently included projects with UNICEF, UNESCO, The World Bank, Save the Children, and the Global Campaign for Education.

We are constantly finding new ways to answer key educational questions to create better opportunities for learners worldwide.

NFER
International

Evidence for excellence in education

P