Introduction and background

In 2015, the United Nations General Assembly laid out 17 global goals for sustainable development. Goal 4 calls for global efforts to achieve inclusive quality education for all. In order to achieve this, there is a need to understand teaching and learning as it happens in the classroom. This is particularly the case in Low and Middle Income Countries (LMICs) where education faces several contextual challenges. As a consequence, classroom observation tools have gained increasing prominence.

This guide from the National Foundation for Educational Research (NFER) evaluates 3 widely used classroom observation tools: The Stallings Classroom Observation System (Stallings), the Classroom Assessment Scoring System (CLASS), and Teach. The NFER’s evaluation was supplemented by 15 interviews with experts who have experience in the development, evaluation, and implementation of classroom observation tools. The full report provides examples of how the tools have been used across a range of contexts.

Key points

The value of observation tools

- Observation tools can capture 3 key elements of classroom practice. Firstly, they look at how teachers teach, answering questions about teachers’ motivations, skills, behaviour management, engagement, and the support and resources needed for success. Secondly, understanding classroom practice can aid inclusion by shedding light on the extent to which teachers are willing and able to include all students in the lesson. It can identify the support which they may need to do so. Finally, observation tools can capture elements of socio-emotional learning – skills beyond numeracy and literacy which are increasingly recognised as key to students’ long-term success. The development of socio-emotional skills is particularly important for disadvantaged students who come from stressful backgrounds.

The Stallings Classroom Observation System

- Stallings was developed by Professor Jane Stallings in 1970 to research the quality of basic education in the US.
- Stallings looks at 4 variables: teachers’ use of instructional time; teachers’ use of materials, including Information Communication Technology; core pedagogical practices; and teachers’ ability to keep students engaged.
- 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 materials the teacher is using, how many students are engaged, and what those who are not engaged are doing.
- Results can be measured against benchmark targets for various activities. For example, the best practice benchmark for instruction (active and passive) is 85%, and the benchmark for student off task behaviour is 6 per cent or less.
- The main benefit of Stallings is its simplicity. It generates rigorous and comparable quantitative data. Results are comparable across schools, regions, or countries, and can be used to develop indicators that benchmark and track progress over time. Findings can be easily communicated to inform and influence educational reforms and priorities. Stallings is cost effective. It 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.
- Stallings should therefore be used to generate rigorous ‘snapshots’ of what is happening in the classroom, when easily communicable data are required, or when there is a need to conduct large scale research across a range of different contexts.
- Stallings cannot capture the variables which may affect a teacher’s performance from day to day, or with different classes. It is therefore not recommended as a high stakes individual performance assessment tool. Stallings cannot measure the quality of teaching content itself. Two teachers may appear very similarly in Stallings metrics but be delivering teaching of qualitatively different levels due to the nature of their interactions with students. Furthermore, it is not always sufficient by itself to explain student learning outcomes. For example, teachers in Afghanistan have been found to have low absence rates and relatively high time on learning activities, but because of their very low content and pedagogical knowledge, students are not necessarily learning. Stallings does not (and was not designed to) provide insights on gender, disability, and socio-emotional learning.

The Classroom Assessment Scoring System

- CLASS (Classroom Assessment Scoring System) was initially developed in the USA and has since been implemented in over 50 countries.
• 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.

• CLASS explores 3 broad domains. The first is emotional support. This domain 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, with their ideas and opinions respected and valued.

• The second is classroom management. This domain 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 can manage time and routines to maximise the instructional time. The third domain, instructional support, examines 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.

• The 3 domains 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 CLASS system is complex, and researchers must be fully trained. 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.

• CLASS explores how teaching processes drive learning. It provides greater depth of information around teacher quality and may have greater potential to explain the link between teaching practices and learning outcomes. Research 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. CLASS is a particularly useful tool for identifying and providing teachers with individualised and tangible feedback to improve their teaching instruction. CLASS has been adapted for use with a wide range of age groups, from infant through to secondary level.

• CLASS is a trademarked tool and is not open source, therefore there are greater costs associated with accessing and delivering the observations. Observers are required to undergo paid training and certification. They need to be skilled and experienced. In certain contexts, teachers may have limited ability to respond to feedback.

Teach

• 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.

• Teach was designed for use at primary level, and in 2021 was expanded to early childhood education. At the time of writing, it is being developed for use in secondary school.

• Teach Primary tracks the amount of time that teachers and students spent on learning activities, and the quality of teaching practices as explored through 3 domains. The domains are classroom culture (positive learning environments), instruction (methods which deepen understanding and encourage analysis), and socio-emotional skills.

• Teach measures 2 main components of teacher effectiveness. Time on task observation involves 3 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). They need to be skilled and experienced. In certain contexts, teachers may have limited ability to respond to feedback.

The full documents can be downloaded from:

https://www.nfer.ac.uk/media/5102/understanding_the_classroom_a_guide_to_selecting_classroom_observation_tools.pdf