

Classroom Observation Tool

**Methodology
2021**

Published January 2022

Table of Contents

| | |
|--|-----------|
| CLASSROOM OBSERVATION TOOL OVERVIEW | 3 |
| TOOL DESCRIPTION | 3 |
| TOOL USES | 3 |
| TOOL DEVELOPMENT | 4 |
| TOOL APPROACH | 4 |
| TOOL DEVELOPMENT PROCESS | 4 |
| PILOT TESTING STAGE | 5 |
| FIELD TEAM TRAINING | 8 |
| COB TOOLKIT | 8 |
| INTER-RATER RELIABILITY | 8 |
| DATA QUALITY CONTROL AND ANALYSIS | 9 |
| ONGOING QUALITY CONTROL | 9 |
| DATA ANALYSIS AND INTERPRETATION | 9 |
| CONCLUSION | 10 |
| LIMITATIONS | 10 |
| FUTURE DIRECTION | 10 |

Classroom Observation Tool Overview

Tool description

The Classroom Observation (COB) Tool is a bespoke tool that aims to gather valid and reliable data on the in-class teaching practices, student engagement and classroom environment at Opportunity EduFinance partner schools. Partner schools are defined as affordable non-state schools that choose to enrol in EduQuality, a three-year holistic school development program offered by Opportunity EduFinance. The objectivity and quality of the data is maximized by using trained external Measurement & Evaluation (M&E) Specialists to carry out the observations, rather than relying on teachers' self-report measures or observations by school leaders. The COB tool consists of several portions:

- **Background Information.** Lesson subject and grade, as well as scheduled and actual start and end times of the lesson, are recorded in the background information section.
- **Learner Engagement.** This section uses a time-interval ('snapshot') approach to do a 10-second scan of the classroom four times at five-minute intervals, and record whether a learning activity is taking place (as opposed to activities unrelated to learning), and if so, how many learners are off-task.
- **Materials and Environment.** A checklist approach is taken to indicate if certain key educational materials are available in the classroom, and if so, whether they are used throughout the lesson.
- **Planning and Record-keeping.** Likewise using a checklist approach, a record is taken of whether the teacher has a plan for the observed lesson, and if so, whether the plan includes learner-centred activities and assessments. It is also noted whether the teacher has a marking record (that is, a system for recording students' marks across time).
- **Teaching Practice.** This section of the tool records the quality of teaching practice observed throughout the lesson on seven internationally recognized and EduQuality programme-aligned indicators of learner-centred teaching. Two timed indicators are also included, recording the number of minutes dedicated throughout the lesson to teacher lecture and/or student discussion, respectively.

Tool uses

Overall, the COB tool is designed to assess the availability and quality of student-centred teaching behaviours, classroom environments and lesson preparation practices in EduFinance partner schools. Several purposes and uses for this tool are envisioned:

- **Donor accountability**
- Evaluation of **effects of the EduQuality programme** (and the Teacher Mentor Professional Development component specifically) on classroom practices across time
- **Feedback to programme management** about areas of teachers' growth and continued need for support/ training (formative function)

Tool Development

Tool approach

COB is a customized tool designed for Opportunity EduFinance based on several internationally recognized classroom/ teacher observation tools and empirically based best practices. While an existing observation tool could have been used, key factors contributing to the decision of creating a bespoke tool included:

- Alignment with a variety of key Opportunity international **donor indicators**;
- Alignment with **typical contexts** of EduQuality's work and **characteristics of the schools** with which it partners;
- Alignment with **EduQuality** guidance and training programmes;
- Inclusion of **different types of indicators**: both teaching practices and other lesson aspects (classroom environment, lesson planning etc.); indicators of time, availability and quality;
- Ability to use the tool **across grades** (pre-primary, primary and secondary school),
- Ability to use the tool **across subjects** (literacy/ language, numeracy/ mathematics, science),
- **Capacity** of national M&E teams to make reliable judgments on tool indicators;
- **Scalability** (i.e., minimal training/ pedagogical expertise required to make reliable judgments on rigorous indicators across time and markets, as well as easily analysable data).

Tool Development process

Keeping in mind the above considerations, the tool development team used established observation tools – in particular the World Bank's Teach¹ tool – as the foundation for internationally-recognized standards in teaching and a validated tool approach on which to base the COB tool. They were further adapted per EduFinance's needs and context, as described below.

| COB section | Approach to section design |
|--------------------------------------|--|
| Learner Engagement | A combination of the Stallings ² and Teach tools' time-interval sampling approach was used to measure time on task, as well as learner engagement during learning activities. COB uses 4 snapshots to measure time on task throughout the lesson, while Stallings and Teach use 10 and 3 snapshots, respectively. Four was judged to be the highest feasible number of snapshots for COB, considering the observers' capacity, and in light of simultaneous observation of the other elements of the tool. |
| Materials & Environment | These two sections of COB follow a simple checklist approach. The items were selected based on the context of EduFinance's partner schools (e.g., gathering information on whether textbooks are available), as well as EduQuality intervention emphasis areas (e.g., student-centred lesson planning). |
| Planning & Record-keeping | |
| Teaching Practices | The teaching practices chosen for this portion of the COB tool were initially based on international best practice of effective teaching approaches, as reflected in the Teach tool and wide-ranging empirical evidence. |

¹ www.worldbank.org/education/teach

² <https://documents1.worldbank.org/curated/en/790221467997639302/pdf/97904-WP-Box391498B-PUBLIC-WB-Stallings-web.pdf>

| | |
|--|---|
| | <p>An alignment exercise was then carried out, whereby the extensive compiled list of key student-centred teaching approaches was compared to the teaching practices encouraged by the EduQuality programme (through Pathways to Excellence guidance and Teacher Mentorship training). The teaching practices list for COB was refined and finalized to match the EduQuality intervention, also based on international best practice.</p> <p>As the COB tool is to be used across subjects, its teaching practice items do not assess presence or quality of specific teaching techniques (e.g., “I do – we do – you do”, “turn and talk” etc.), or subject-specific strategies, but rather general adherence to pedagogical best practice, which can be achieved using a variety of techniques and in any academic context.</p> <p>The Teach scale of Low-Medium-High was adopted for the Teaching Practices section, with a score of None added to account for instances where a teaching behaviour does not occur at all. During the tool pilot-testing stage, some of the indicators were judged not to reliably lend themselves to a four-point scale (None to High) and were instead transformed into binary (Yes/No) indicators.</p> <p>Two of the Teaching Practices – Lecture and Discussion – could not be assessed using the quality approach and required instead a timing approach. Both are assessed as time-interval indicators in the Stallings tool. Due to only 4 snapshots in the COB tool, compared to 10 in Stallings, a direct timing approach (in minutes), instead of time-interval sampling, was chosen to capture Lecture and Discussion in COB.</p> |
|--|---|

When designing the COB tool, **low-inference items** (readily observable, objective) were preferred to **high-inference items** (requiring subjective qualitative judgments), when possible. Nevertheless, some high-inference items were selected for the Teaching Practices section specifically. A combination of both item types provides a broad-stroke overview of lessons and classrooms in EduFinance partner schools (e.g., availability and use of certain materials), as well as an in-depth look at teaching quality. Continuous variables (e.g., 1-4 scale and time in minutes) also allow for more granular tracking of incremental growth across time.

During the tool development phase, the COB design team attended a Global Reading Network³ workshop on creating customized Classroom Observation measures. Best practice on bespoke instrument development shared during the workshop and in its corresponding published guidance⁴ influenced the design of the COB tool for Opportunity EduFinance.

Pilot testing stage

COB tool drafts went through multiple initial iterations as a result of internal discussions and revisions. Education Specialists who work in the local contexts reviewed and gave feedback on the tool until face validity was reached, with agreement between a diverse group of stakeholders on the tool’s intended meaning (international consultant, EduQuality programme management, data team management, field education specialist teams, M&E teams).

³ <https://www.globalreadingnetwork.net>

⁴ Hertz, A. C., Kochetkova, E. & Pflepsen, A. (2019). *Classroom Observation Toolkit for Early Grade Reading Improvement. A Global Reading Network Resource*. Prepared by University Research Co., LLC. (URC) under the Reading within REACH initiative for USAID’s Building Evidence and Supporting Innovation to Improve Primary Grade Assistance for the Office of Education (E3/ED).

This was followed by pilot-testing the COB tool while watching video-recorded lessons. The tool design team and field managers independently scored these videos, then compared and discussed scores, adapting the tool for areas of frequent disagreement. This process led to changes that made the tool more user-friendly and reliable. For example, the team eliminated hard-to-measure indicators (such as the initial intent to time the duration of every teaching practice), added some useful indicators (e.g., split hands-on materials use between the teacher and students, rather than keeping it combined as one indicator) and changed some indicators into easier-to-measure types (e.g. due to lack of inter-rater⁵ agreement on a quality-rating scale, some indicators were changed to binary ones). At other times, despite disagreement on scores between raters and the master-rater who designed the tool, indicators were kept, but written guidance and training on assigning scores on it were strengthened.

As a result of this iterative process, the inter-rater agreement during the pilot stage reached 80% and above, leading to the judgment that the COB is a reliable tool that is ready to use in the field. A brief analysis of the tool's readiness for use is presented below, using a checklist from the Classroom Observation Toolkit⁶.

Tool validity and reliability

| Tool quality questions | Responses |
|-------------------------------------|--|
| Is the instrument VALID ? | <p>Yes. COB was designed by an international expert in general education and measurement. The instrument was based on internationally validated tools and aligned with best practice (teaching strategies and lesson characteristics that are empirically demonstrated to lead to better learning outcomes). Further, COB was reviewed by field managers and staff, as well as data managers, for appropriateness to context and data analysis capacity and needs.</p> |
| Is the instrument RELIABLE ? | <p>Yes. The COB tool has undergone several iterations, each version increasing in reliability, established through:</p> <ul style="list-style-type: none"> ○ Agreed-upon and recorded operational definitions of all indicators in the Coding Manual, ○ Extensive practice scoring of video lessons by several observers, resulting in items with constantly low inter-rater agreement being eliminated, changed or specified more clearly in the Coding Manual, ○ An average inter-rater reliability (IRR) of at least 80% across items for the data collection team was expected before the instrument was considered reliable and specialists ready for field data collection (individual IRR was also tracked, and M&E specialists not considered ready for independent data collection until their IRR with the master-scorer was 80%). |

⁵ "Inter-Rater Reliability"(IRR) Is the extent of agreement on scores between different observers. In this case, IRR was calculated by comparing each observer's score to the score assigned by the master-scorer (who was also the tool designer). The percentage of scores on which each observer agreed with the master-scorer was considered that observer's IRR score. A consistent IRR of 80% or above signaled that the observer was ready for independent observation.

⁶ Ibid.

| | |
|---|--|
| Does the FOCUS of the instrument align with what it aims to measure? | Yes. The focus of the instrument was determined by empirically based best practice in teaching, as well as areas of emphasis in the EduQuality programme. As a result, the tool will be used to measure teachers' growth across time in effective teaching practices, aligned with the training and guidance they and school leaders receive from EduQuality. |
| Is the Item and response FORMAT appropriate? | Yes. Considering data-collection team expertise and capacity, low-inference items are used where possible (snapshots, yes/no binary items, timing lecture & discussion). Of 63 indicators in the main portion of the COB tool, 59 are low-inference items (94%). A few high-inference items (quality judgment) were included to provide greater depth of data, more nuanced information for programme management and a more granular scale for long-term growth tracking. Data collection for these items was made possible through in-depth descriptions/ definitions of the items and their scoring in tool protocol documents, as well as extensive observer training and practice. |
| Do the INSTRUCTIONAL CATEGORIES and BEHAVIORS align with the purpose and the instructional strategies that the program is interesting in observing? | Yes. The COB tool is aligned with teaching-related EduQuality programme components (Pathways to Excellence and Teacher Mentor training), observing lesson elements that are consistent with the EduQuality intervention. All items are also based on evidence-based teaching best practice. ⁷ |
| Is the instrument FEASIBLE to implement? | Yes. The COB tool was designed with local contexts and observer capacity in mind. Further, it has been field-tested and found feasible to implement in the span of one lesson (including making final judgments after the lesson's end). A variety of item types add to the instrument's complexity, which is addressed by providing extensive observer training before data collection. The combination of item types was modelled after the internationally recognised Teach Tool. |
| Is the instrument USER-FRIENDLY ? | Yes. The Collect digital application ⁸ , which is familiar to all EduFinance M&E specialists, is used to collect data using the COB tool. Further, a guided Internal Notes Template has been developed for observers to take guided notes to facilitate judgments on high-inference indicators. |

⁷ Pathways to Excellence and Teacher Mentor training are likewise evidence-based.

⁸ Collect is an application provided by Atlan (<https://atlan.com/>)

Field Team Training

COB toolkit

To support observers' understanding of the COB items and ensure reliability, a kit of materials with instructions was developed to accompany the tool, based considerably on the Teach approach to such guidance and adapted according to partner schools' context and field pilot testing. The first portion of the M&E Specialist (observer) training focuses on establishing a thorough understanding of the COB using the materials in the toolkit. Explanations by trainers, ongoing quizzes and Q&A sessions are designed for observers to arrive at a similar conceptual understanding of the tool.

The COB toolkit consists of:

- **Coding Manual:** operational definitions of every COB item; descriptions and examples of all scores on each item; methods of scoring and notes to correct common mistakes in scoring each item.
- **Data Collection Protocol:** guidance on the logistics and ethics of the school visits and in-class observations; considerations of ensuring objectivity when scoring.
- **Internal Notes Template:** guided note-taking sheets with summaries and key reminders about each high-inference item (Teaching Practice).
- **Animated Demo's:** micro-videos illustrating examples of teaching practice consistent with each score on high-inference items (Teaching Practice).

The tool training has been digitised, with all sessions, presentations and quizzes recorded, reducing future training costs. A blended-learning approach will be deployed moving forward, as new M&E specialists join the observer team and/or new markets are added to the EduQuality program.

Inter-rater reliability

The second step in observer training is extensive practice scoring video recordings of lessons from contexts similar to those of EduFinance partner schools. Observers are asked to score 6-8 practice videos, submitting their COB scores using the Collect platform after each one, followed by a joint review session where correct scores and reasons for them are discussed. An M&E specialist is considered ready to independently collect COB data in the field when they have reached 80% agreement with the master scorer across all items. It is also important that the whole cohort reaches 80% agreement with the master scorer on each COB item, ensuring that scores across the items are equally reliable.

Data Quality Control and Analysis

Ongoing quality control

To further ensure reliability of the data, individual check-ins with M&E specialists are scheduled after the first 2 weeks of their data collection, to discuss areas of doubt or questions. Their open-ended notes for each high-inference item (captured in the Internal Notes Template) are compared to the scores they assigned. In cases of misjudgements, scores are changed and reasons for the misjudgement discussed, as part of observer capacity building. Moreover, for observers with lower IRRs at the training stage, joint independent observations with a partner are practiced in order to check areas of alignment between them.

Beyond the initial observation stage, individual check-ins take place periodically, with local M&E associates undertaking ongoing data validation by comparing observers' scores to their qualitative notes.

Data analysis and interpretation

Two broad approaches can be taken to COB data analysis and interpretation:

1. Per-indicator analysis

Each indicator can be analysed to give a granular portrait of the teaching practices and classroom environments in EduFinance partner schools. This type of analysis is likely to be most useful for the purpose of informing programmatic interventions, such as the teacher mentor professional development sessions, or the school leadership professional development sessions.

When indicators are taken together or minor calculations are performed, bigger-picture questions can be answered with the data, such as:

- Do teachers begin and finish lessons on-time?
- What percentage of instructional time is dedicated to learning activities?
- What percentage of students are on-task during learning activities?
- Are learning aides and essential teaching & learning materials available to teachers and students?
- Do teachers use materials available to them, during lessons?
- Do teachers plan learner-centred lessons?
- Do teachers have a system of recording students' marks across time?
- What proportion of instructional time is dedicated to teacher-centred lecture?
- What proportion of instructional time is dedicated to learner-centred discussion?
- How deep and conceptual are teachers' explanations and students' active learning tasks?
- What proportion of the students do teachers check for understanding throughout the lesson?
- Are lessons connected to the broader subject curriculum?
- Do teachers give clear and specific instructions for students' tasks and/or expected behaviour?
- Do teachers make intentional connections between the lessons and students' real-life experiences?
- How supportive is the classroom culture fostered by teachers?

2. COB Composite score

A composite score can also be used, particularly when the purpose is reporting on or evaluation of overall lesson quality or growth over time, rather than analysis of its sub-components. While this approach is challenging and more subjective than the per-indicator approach, it is easy for interpretation and making big-picture conclusions.

Based on informed, well-founded decisions, one score can be built from all or some of the COB indicators. In order to do this, weights are assigned to each COB indicator, proportionately to its comparative importance to or effect on the overall quality of the lesson. Subsequently, scores on various types of indicators are translated into one scale, so that binary yes-no scores, scores on a scale of 1-4, percentages of learners off-task and scores of time spent on lecture or discussion are aligned to one standard of measurement. Finally, a weighted average of scores across the chosen indicators is calculated, resulting in a single composite classroom observation score.

Conclusion

Limitations

As is the case with every tool, the COB has some limitations. First, being an observation tool, the COB only captures visible teacher behaviours and classroom elements – it cannot assess teachers' intentions. For example, the tool cannot make a distinction between whether a teacher has carried out an assessment during the lesson for summative or formative purposes. The tool only captures how many learners' answers or work has been checked during the class. This measure approximates formative assessment, but only in a limited way. Thus, interpretation of findings needs to be done carefully, keeping this in mind.

Second, capacity and expertise of local M&E teams is often limited, presenting a risk to reliability, particularly of high-inference items. For this reason, the initial COB data-collection training program is intensive, the supporting toolkit is detailed, and ongoing data validation is envisioned throughout the life of the project.

Third, some of the indicators are subjective by nature and require qualitative judgment. While data on these indicators are, as a result, at risk of being less reliable, these measures are considered important to include. The combination of more objective, low-inference items and these richer, qualitative measures creates a more useful and detailed picture of classroom practice.

Future direction

It is envisioned that the COB data will be used long-term to evaluate the effectiveness of EduQuality's Teacher Mentor Professional Development sessions, as well as other relevant interventions, on changing practices in the classroom. This will in turn be reported to donors and feed back into program development. In the long run, it is hoped that the rigorous process of collecting bespoke data on the quality of classroom practices at EduFinance partner schools will lead to a better-informed program that fits the areas of teachers' greatest need. Subsequently, the program will result in growth in teacher skill levels, to be captured yet again in the COB data, continuing the data-driven program improvement cycle.