The Measurement Crisis: An Assessment of How Countries Measure Classroom Practices

Julia Ladics  
*Brown University*

Ezequiel Molina  
*The World Bank*

Tracy Wilichowski  
*The World Bank*

Noah Yarrow  
*The World Bank*

March 2018

**Abstract**

Given the importance of diagnosing teacher strengths and weaknesses to inform professional development, this research seeks to uncover how many countries use observation tools, what these tools look like, whether they’re evidence-based, and the extent to which they’re reliable. This paper is the first of its kind to evaluate how countries currently measure classroom observation practices from a global dataset of thirty developing countries.
INTRODUCTION

After several years of schooling, students are not acquiring basic foundational knowledge -- a phenomenon that has been dubbed "the learning crisis." In an effort to understand why children are not learning in developing countries, a recent seven-country study collected evidence on the quality of service delivery in Sub-Saharan African schools and found high levels of absenteeism, low content and pedagogical knowledge, and ineffective teaching practices (Bold et al. 2017). Even teachers who were consistently in the classroom had extremely low content and pedagogical knowledge. For example, almost a quarter of teachers surveyed could not subtract double-digit numbers and one-third were unable to multiply double-digit numbers. Poor content knowledge was mirrored by low-quality instruction. On average, less than one in ten used effective teaching strategies, such as lesson structuring, planning, asking lower- or higher-order questions and providing feedback, to promote learning.

In an attempt to improve learning at scale, teacher training programs have been deployed around the world. However, as the World Bank’s World Development Report 2018 notes, most of these programs are highly ineffective. Some of the problems include a lack of customized support, which is not tailored to the needs of the teacher, no follow up visits after the training, and no link between the training and salary or performance. Moreover, these training programs are often criticized for being largely theoretical, without providing opportunities for teachers to apply what they learn.

To close the gap between what works and what is implemented in teacher training programs, it is essential to accurately measure what teachers do in the classroom. If teachers’ strengths and weaknesses are not identified, they cannot be addressed, and their teaching practice is unlikely to improve. The question is then why are classroom observation tools needed to measure teacher weaknesses and strengths? Can’t teachers and principals simply gauge the quality of instruction, given their experience? The answer is no. In fact, a recent study found less than 50% of teachers and principals were able to identify ‘effective’ teachers as measured by value added scores, which is akin to chance (Strong et al. 2011). That is to say, when asked to measure good or bad instruction, they only got 1% of those judgments correct (for poor instruction) and 4% correct (for good instruction); overall, this means 63% of the judgments did not reflect the quality of instruction observed.

This paper details the percentage of classroom observation tools that reliably measure teaching practices tied to student outcomes and provides recommendations for countries that are currently not.
METHODOLOGY

Developing a Framework
To analyze the content of these observation tools, a conceptual framework is used, which is broken down into areas, elements, and behaviors. This framework was developed by collecting an inventory of elements from commonly used international tools. A study by Gill et al. (2016) was also used to inform the conceptual framework, as it compared five widely used international classroom observation tools and analyzed their content. The tools used in this study were sorted into common areas and elements to which additional areas and elements from international tools were added.

Second, a working group of thirty education specialists and economists were convened to examine the elements in this inventory, and narrowed the list. To facilitate the creation of this inventory, the working group discussed three questions: what effective teachers do, what can be measured in a classroom observation tool, and what is relevant for policy makers.

Lastly, because stages one and two are relevant to international tools, two additional steps were added to add relevancy to the final list of national tools. First, each national tool was reviewed for elements missing from the inventory. Second, two rounds of coding were conducted. Between the rounds, elements from national tools that were not captured in the initial round were noted and included in the final list of elements.

Coding Methodology
In order to capture the variety of material present in the sample, concepts were broken down into three areas, 30 elements, and 63 behaviors. Elements and behaviors were coded on a scale from zero to two depending on the nature of measurement. Zero was used for tools without elements. One was used if there are elements but they are recorded using check boxes or yes/no questions. One also captured elements recorded as ratings, in the absence of a rubric or examples to guide the observer when making rating decisions. A two was used if the element exists in the tool and is recorded using a ratings scale, for which the observer has to refer to a rubric or example behaviors, or one in which key terms are clearly defined.

The key difference between a one and a two is how much information is provided to the observer to decide on the code. Another point of differentiation is how much information the tool collects, given the way the questions are phrased and answers are recorded. Tools that record answers using check boxes or yes or no questions are coded as one, and are understood to measure elements broadly. Conversely, tools that record answers as ratings and have a complementary rubric with specific examples, are coded as two, and are understood to measure elements specifically.  

---

1 See Appendix A for a full list of areas and elements that includes behaviors.
2 See Appendix B for a detailed explanation of the coding process.
What is Evidence-Based?
To understand which elements are evidence-based, evidence was collected using three systematic reviews and twelve research papers published between 2005 and 2016. The twelve research papers alone include a wide range of evidence from the United States, Switzerland, Ghana, Pakistan, India, Honduras, Colombia, Mexico, Brazil, and Chile. The review papers draw results from an even wider sample.

To avoid mis-categorizing elements, they were not sorted into those that do or do not predict student outcomes. Instead, 15 of the 30 elements were put into three categories depending on the degree with which there is evidence they predict student outcomes. The remaining 15 elements are descriptive, and are left as unrated. Behaviors are not sorted into these categories, as they are much more difficult to accurately map to the conceptual review framework.

Table 1: Elements Relationship to Student Outcomes

<table>
<thead>
<tr>
<th>Low Relationship</th>
<th>Mixed Relationship</th>
<th>High Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements have a low relationship to student outcomes if none of the papers reviewed mention they have a positive relationship to student outcomes, or if the papers indicate they have a negative relationship with student outcomes.</td>
<td>Elements have a mixed relationship to student outcomes if at least one paper has found mixed evidence from the same country about the relationship between the element and student outcomes. For example, the element may have a positive relationship with high-income students’ outcomes, but no relationship to low-income students’ outcomes. Alternatively, elements may have a mixed relationship to student outcomes if there are an almost equal number of papers in which the authors find a positive relationship, and a zero or negative relationship, between the element and student outcomes.</td>
<td>Elements have a high relationship to student outcomes if multiple papers have found evidence of a positive relationship between the element and student outcomes, and none have found negative relationships. Alternatively, elements have a high relationship to student outcomes if at least one paper has found consistently positive relationships between the element and student outcomes across multiple student demographics.</td>
</tr>
</tbody>
</table>

Description of Elements in this Category

3 See reference section for full bibliography.
4 Note: All other elements are unrated. These are largely descriptive elements, or elements that neither predict student outcomes nor are expected to predict student outcomes. They are not included in the analysis of the degree to which tools are evidence-based.
<table>
<thead>
<tr>
<th>Examination of classroom environment</th>
<th>Classroom management</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality and inclusiveness</td>
<td>Critical thinking</td>
<td>Check for student understanding</td>
</tr>
<tr>
<td>Family engagement</td>
<td>Student focus (driven entirely by teacher listens actively and encourages ideas)</td>
<td>Content understanding</td>
</tr>
<tr>
<td>Professionalism and respect</td>
<td>Social climate</td>
<td>Clear presentation and lesson structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Language development</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lesson facilitation and discourse</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Motivating the classroom (driven entirely by communicating high expectations)</td>
</tr>
</tbody>
</table>

**Quantifying Reliability**

In order for tools to accurately identify teachers’ strengths and weaknesses, observers must be held to a certain reliability standard, meaning, they must be trained and assessed before observing classrooms. Reliability refers to the idea that any two observers who view the same lesson, would independently assess the classroom similarly, most of the time.

Given Strong et al.’s (2011) research findings. It’s extremely important for observers to be reliably trained. As such, the following four questions were posed to examine the reliability of each tool in the sample.

1. Do the observers need to undergo any training to use the tool? How long is the training?
2. Did the government or another agency measure the reliability of observers using the tool?
3. Is there any evidence or documentation that inter-rater reliability was calculated; if so, what methodology was used?
4. Are observers required to get certified, meaning did they have to pass a reliability exam with minimum thresholds to use the tool?

The answers to each of the four questions were recorded to compute the extent to which the tools in the sample have a reliability protocol.

**Sample**

The final sample consists of 39 primary school observation tools from 30 countries. Four international tools were also included as a benchmark for comparison (see Figure 1 for a visual representation). The sample includes two tools from Central and Southern Africa, two tools from West and Central Africa, ten tools from East Africa, two tools from the Middle East and North Africa, three tools from Europe and Central Asia, five tools from South Asia, six tools from East Asia and the Pacific, and nine tools from Latin America. There are no national tools from North America or Western Europe. For the current version of this paper, 20 observation tools are used as the 10 non-English tools are still being translated and processed.

**Figure 1:** Map of countries in the entire sample
Note: Red indicates countries for which there are tools in the sample. Blue indicates other countries that have currently have active projects and we attempt to obtains tools from.

The Tool's Purpose
Observation tools can be used for many purposes, such as professional development, inspections, and as measures of accountability. Although there is limited information on the purpose of the tools in the sample, inferences were drawn based on the coding process and content from the tool’s supporting materials.

Most of the tools in the sample appear to be used, in part, for professional development. For example, sixteen of the tools ask observers to take notes at the end of the observation – over half provide either recommendations or goal-setting for the teacher, and six require the observer to discuss the findings with the teacher after the observation. Two of the remaining four use rating scales with a rubric, which implies they are also likely to be used for professional development. With this information, it’s assumed 90% of tools in the sample are used as mechanisms of professional development.

With this said, approximately one third of the sample seems to be used for accountability. Seven tools in the sample grade the teacher for each of the tool’s categories, while five provide a final grade at the end of the observation. Ten tools also require either a discussion with some students about the lesson content, or an examination of student work after the observation. Similarly, one third of the tools are used to measure the implementation of a program. Seven tools clearly indicate they measure the implementation of a specific program, and therefore ask targeted questions and collect key implementation information. Nevertheless, both the tools that measure program implementation, and those that are used for accountability, have aspects of professional development.
DATA ANALYSIS

The analysis seeks to answer the four questions laid out in the introduction: (i) What percentage of World Bank client countries use classroom observation tools?, (ii) Of these, what does the average tool look like?, (iii) Are the tools evidenced-based?, and (iv) Are the tools reliable?

I. Are Countries Using Classroom Observation Tools?

The World Bank’s Education Practice has 144 active education projects, which encompass 80 countries. The sample used in this dataset was collected by reaching out to task team leaders (TTLs) for these 144 projects and incentivizing them to send classrooms observation tools. They submitted classroom observation tools for the countries in which they work – this was contingent upon the existence of such tools and the appropriate government permissions. Over 60 submissions were received; however, not all 60 were applicable to the study. Of the initial sample, we eliminated tools used for randomized controlled trials (RCTs), early childhood education, higher education, and school inspection frameworks were removed from the sample.

The final sample consists of 39 primary school observation tools from 30 countries. 4 international tools were also included as a benchmark for comparison. It’s estimated 37% of countries with active portfolios at the World Bank use classroom observation tools at primary school level. Because we cannot be sure we got all the tools available we estimated, using a back of the envelope calculation, that at most 50% of the countries the World Bank has active portfolios use a classroom observation tool in their education system. Of the countries that do collect primary school classroom dynamics, only a small percentage measure them regularly. Most tools in the sample are either recently developed, or had been used for a short timespan (two or three years) and have not been used since.

II. The Average Tool: What are Countries Measuring?

Of particular interest is of the 50% of countries that have classroom observation tools, what do the tools look like? The “average tool” is defined both by the elements it measures and the make-up of the instrument.

Some of these elements include collecting data and checking the classroom environment, usually at the beginning of the lesson. For example, a tool may count the number of girls and boys in the classroom, and note absences, noting which of those students have textbooks or other materials. Tools also measure general teaching related elements; such as clearly presenting content, managing the classroom, and facilitating the lesson. They may check whether the teacher outlines clear learning objectives, manages time well, and uses different types of teaching techniques. Some tools check if the teacher employs specific pedagogical actions, such as checking for understanding, providing feedback, and creating a positive social climate.

At the end of the observation, the tools often assess teachers’ ability, provide recommendations, or require observers to take a series of notes. The average tool records these responses with a checklist or yes/no questions. Twelve tools use a checklist or yes/no questions, while eight tools use a rating scale, three of which have a rubric or example behaviors and definitions.

---

5 http://projects.worldbank.org/search?lang=en&searchTerm=&mjsectorcode_exact=EX
III. Are Tools Evidence-Based?

In addition to what tools are measuring, it’s important to assess how many of the countries classroom observation protocol have elements that are related to student outcomes. These tools must also be examined to assess the extent to which these elements are captured in a broad or specific manner. After assessing this dataset, it’s found that over 80% of the elements from tool’s in our sample are likely to have at least a mixed relationship to student outcomes. However, while tools measure some key behaviors that are closely related to student outcomes, they do not measure any element consistently and specifically. This means that do not define the words or provide examples. On average, only 33% of elements are measured specifically. For elements that appear in at least half of the tools in our sample, this number falls to only 25%.

Distribution of Element Types

In addition to measuring whether the tools capture specific elements, the study also assessed whether the elements in the sample have a low, a mixed, and a high relationship to student outcomes. This measurement aims to show the extent to which classroom observation protocols in the sample have each type of element, and subsequently implies the degree to which tools are evidence-based.
As figure 3 shows, on average, 19% of tools measure elements with a low relationship to student outcomes, 25% of tools measure elements with a mixed relationship to student outcomes, and 56% of tools measure elements with a strong relationship to student outcomes. To understand the variation in the sample, the tool with the lowest percentage of low ranked elements, and the tool with the highest percentage of low ranked elements is also emphasized. The tool with the lowest percentage of low ranked elements contains 17% low and 83% high elements, while the tool with the highest percentage of low ranked elements contains 50% low, and 25% high and mixed.\(^6\)

**Figure 3: Percentage of tools that contain each element category**

For comparative purposes, these exercises were also conducted for the international tools. On average, international tools measure 14% low, 37% mixed, and 49% high elements. While they measure fewer low elements than national tools, they measure fewer high and more mixed elements. There is also more variation between national tools than between international tools.

**IV. Digging Deeper: Which Behaviors Do Tools Measure?**

As discussed previously, parent elements are classified by their relationship to student outcomes. By association, behaviors belonging to these parent elements are treated as equally related to student outcomes. The behaviors are examined to further understand how tools measure each of the three categories of elements.

Tools tend not to measure any of their elements thoroughly, and systematically leave out important behaviors. For example, the tools that measure critical thinking measure whether the teacher asks thinking questions, but not whether the teacher assigns thinking tasks, or whether the students provide thinking tasks.

\(^6\) The variation among tools is calculated using weighted averages. This is done twice, using both linear and quadratic weights. Firstly, weights from one through three were used for low through high elements, respectively. In a second calculation weights of square numbers, 1, 9, and 25 were used for low, mixed, and high elements, respectively. The resulting tools and breakdown of elements are the same in both instances.
thinking answers. Similarly, although almost all tools measure checks for understanding, only a quarter ensure that these are related to the learning objectives. Finally, national tools tend to measure behaviors that focus on the teacher, but not behaviors that evaluate the students or teacher-student interactions. International tools measure the same behaviors national tools do, but they also measure many elements that national tools do not. In general, international tools capture more of the events and interactions in the classroom as a whole, whereas national tools focus on observing the teacher.

V. Are Tools Capturing Specific or Broad Measures?
In addition to understanding the elements countries’ measure, assessing the quality of those measurements is just as relevant. Thus far, no consideration has been made as to whether tools measure elements broadly (i.e. with a checklist, yes/no questions, or ratings with no rubric), or specifically (i.e. using ratings with a rubric or example behaviors). This is important as the way these concepts are measured matters. When a tool measures something as broad as, “Does the teacher provide feedback to students?” it is impossible to tell whether the teacher simply tells the children whether their work is correct, or whether substantive feedback is given that helps the children understand the root of their misunderstanding. While feedback is correlated with student outcomes, research shows it is the quality of the feedback, which sheds light on the root of the misunderstanding, that is most important. Broad measures do not clearly define feedback, or any of the other terms they use.

Figure 3 shows a breakdown of which tools fall into the broad or specific categories – further disaggregated by the three methodological areas. For the tools elements were coded on a scale from zero to 2. It shows that the tools that do measure elements within each of these areas, even if they are related to student outcomes, overwhelmingly do so broadly rather than specifically. On average, only 33% of elements are measured specifically. Inevitably, tools with elements collect significantly less information than those that observe teachers using ratings and a rubric, and measure the elements related to student outcomes in a less comprehensive manner.

Figure 4: Disaggregated Coding, by Area
Table 2: Breakdown of coding by some example elements

<table>
<thead>
<tr>
<th>Element</th>
<th>Broad</th>
<th>Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>83%</td>
<td>17%</td>
</tr>
<tr>
<td>Language Development</td>
<td>86%</td>
<td>14%</td>
</tr>
<tr>
<td>Critical Thinking</td>
<td>67%</td>
<td>33%</td>
</tr>
<tr>
<td>Positive Social Climate</td>
<td>71%</td>
<td>29%</td>
</tr>
</tbody>
</table>

The elements and behaviors that are most often measured using a rubric are those that are infrequently measured, and are likely contained in tools that are more comprehensive than the average tool. For example, 100% of tools that measure the behaviors of social competence and racial inclusiveness have specific. However, there are only two tools that measure the former, and one that measures the latter. Similarly, 67% of tools that measure the behaviors of established routines and thinking tasks do so using ratings and a rubric. However, only 3 tools measure each. Likewise, there are 5 behaviors for which 50% of tools use ratings and a rubric, but 3 of the 5 only exist in two tools. The other two behaviors are measured in 6 of 20 tools each. Therefore, the highest proportion of ratings and rubric use for a somewhat common behavior is 50%.

Conversely, for elements and behaviors that appear in over 75% of tools, the average rate at which we code them as a two is 22%. Even when looking at elements that appear in at least 50% of the tools, the average rate at which they are measured using ratings with a rubric is only 25%.

This section has several important findings. While tools do capture elements that are evidence based, most of them do so measuring them very broadly, like for instance, whether the teacher provides feedback to the student, or whether the teacher check for understanding. There are two problems with these broad measures: (i) those are unable to capture the quality of the practice, meaning that it would be the same if the teacher says correct to the students than if he builds up their understanding or clarify a misconception, (ii) they are hard to code reliably as the questions do not define terms or provide examples. As a result, observers need to judge what feedback or check for understanding means and how to apply it for a particular case. For example, if the teacher asks a question and only two students raised their hand, one of them answer it correctly and then the teacher moves on. Did the teacher check for understanding or not? Without more information observers will struggle to code this. Because of these two reasons, we argue that broad measures are not going to be useful for providing teacher with good feedback to improve their practice.

VI. Are Classroom Observation Tools Reliable?

Our analysis shows that while tools measure elements using checklists and yes-no questions, they do measure factors that predict student outcomes. We now break down whether the data these tools collect is accurate and reliable.

We have reliability data on 11 of the 20 national tools in our sample. Given our information from the World Bank Task Team Leaders (TTLs), 10 of these tools require some kind of training for observers to use the tool. This training varies from as long as two weeks to as short as a few hours within one

---

7 That is measured in more than 15% of tools
day. 7 of these 10 tools have some reliability protocol to ensure the tool itself is reliable. 3 of them have calculated inter-rated reliability (IRR) with the help of the World Bank, 3 with the help of another implementing agency, and one using a national quality assurance agency. However, the TTLs were only able to produce the respective documentation from 3 of these 7 tools to show that they have actually compute the IRR. Finally, only one of these tools require an exam or a specific certification to determine observers’ reliability. In this case, we could not access the certification exam, but took their TTL word for it. The idea is that we are computing the upper bound of reliability.

CONCLUSIONS

The Findings
The findings of this study are crucial: At least 50% of countries in the dataset do not measure classroom practices, 36.8% are either not evidence-based, or have broad measures. The research as shown that while the tools do capture evidence-based elements, most of them do so very broadly, like capturing whether the teacher provides feedback to the student, or whether the teacher checks for understanding. As we discussed above, there are 2 problems with these broad measures: (i) they are unable to capture the quality of the instructional practice, and (ii) they are hard to code reliably as the questions do not define terms or provide examples. As such, these findings are difficult to use to inform policy. What’s more, the 12% that have specific measures, are not reliable. Meaning, it’s unlikely the observers are coming to the same conclusion from observing the same class. Most troubling – only 1.2% have specific measures and are reliable. This means that the large majority of these tools cannot be used to diagnose teacher strengths and weaknesses and improve teacher practice. As a result, professional development systems relying on imperfect tools is destined to fail.

Call to Action: Building Public Goods to Help Countries Improve Measurement
Measuring what happens in the classroom reliably is not something that can easily be done – as this paper shows. In order for countries to improve the way they currently measure teacher practices, the Global Knowledge and Innovation Team at the Education Global Practice of the World Bank has developed a new open source classroom observation tool. Using the “Foundations of Observation” literature, the World Bank developed an open source classroom observation tool aligned to good practices for creating reliable observation tools and training observers. The tool was developed to (i) cater to the needs of World Bank client countries, (ii) circumvent proprietary measures, (iii) adapt to different contexts.

As of March 2018, enumerators have been trained on TEACH in the Philippines and Punjab, Pakistan, with an overall passage rate of 96% on the reliability exam.

The World Bank’s open source classroom observation tool - TEACH - has been tested with video footage from Afghanistan, China, Pakistan, the Philippines, Tanzania, Uruguay, and Vietnam; these findings were used to revise the tool’s overall structure. Additionally, an expert technical meeting was held to further assess the tool’s elements. Based on feedback, the tool was revised and tested with additional video footage to inform the foundations of the tool. As of March 2018, TEACH is currently being piloted in Mozambique, Pakistan, the Philippines, and Uruguay. Concurrently, a group of students from the University of Virginia are

applying TEACH to international video footage from twelve countries, which will be used to assess the tool’s validity and reliability across contexts.

TEACH will be used both a diagnostic and action-oriented instrument in Bank-supported projects with a strong focus on informing pre-service and in-service teacher professional development. Additional materials will be developed for coaches, inspectors, and principals to diagnose teachers’ strengths and weaknesses and provide targeted feedback. This will be accompanied by exemplar videos from which teachers can learn and improve their instructional practice. Coaches will also learn to monitor whether the teacher implements these techniques and provide continuous feedback to facilitate improvement. The goal is to close the loop on poor service delivery in education by shedding light on the instructional practices that aren’t happening in the classroom, training teachers on how to implement these practices, and providing systems with the necessary information to impart large-scale change at a school, district, and national level.
REFERENCES


Barbara Bruns, Soledad De Gregorio, and Sandy Taut, “Measures of Effective Teaching in Developing Countries” (RISE, September 2016).


Daniel Muijs, Leonidas Kyriakides, Greetje van der Werf, Bert Creemers, Helen Timperley & Lorna Earl (2014) State of the art – teacher effectiveness and professional learning, School Effectiveness and
# APPENDIX A
Full list of Areas, Elements, and Behaviors

<table>
<thead>
<tr>
<th>Theme ID</th>
<th>ID</th>
<th>Element/Behavior</th>
<th>Description or Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.10</td>
<td>Time snapshot</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.12</td>
<td>Recording of snapshots</td>
<td>How are snapshots recorded, through ticking off set activities, rating set activities, or free hand notes about activities?</td>
</tr>
<tr>
<td></td>
<td>1.20</td>
<td>Ratings based</td>
<td>Sections are separated by indicator, leaving clear spaces between indicators (referring to some teacher standards document)</td>
</tr>
<tr>
<td></td>
<td>1.21</td>
<td>Sections separated by indicator category</td>
<td>Sections are separated by indicator, leaving clear spaces between indicators (referring to some teacher standards document)</td>
</tr>
<tr>
<td></td>
<td>1.23</td>
<td>Has Rubric</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.30</td>
<td>Yes/No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.31</td>
<td>Sections separated by indicator category</td>
<td>Sections are separated by indicator, leaving clear spaces between indicators (referring to some teacher standards document)</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>Split into before, during, after</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.60</td>
<td>Instructions/Indicators directly on observation sheet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.80</td>
<td>Includes head teacher in some capacity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.90</td>
<td>Subject specific</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.10</td>
<td>Data collected before observation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.11</td>
<td>Student gender breakdown</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.12</td>
<td>Student absences</td>
<td>Records number of girls/boys, or ratio between them</td>
</tr>
<tr>
<td></td>
<td>2.13</td>
<td>Teacher’s presence/punctuality</td>
<td>Records how many children absent</td>
</tr>
<tr>
<td></td>
<td>2.14</td>
<td>Teacher’s gender</td>
<td>Records teacher's presence/timeliness</td>
</tr>
<tr>
<td></td>
<td>2.15</td>
<td>Teacher’s education</td>
<td>Records teacher's gender</td>
</tr>
<tr>
<td></td>
<td>2.16</td>
<td>Teacher’s experience (in years)</td>
<td>Notes teacher's education level, either by teacher or by interviewer</td>
</tr>
<tr>
<td></td>
<td>2.20</td>
<td>Materials examined before observation</td>
<td>Notes teacher's years of teaching experience either by teacher or by interviewer</td>
</tr>
<tr>
<td></td>
<td>2.21</td>
<td>Teacher’s lesson plan</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.22</td>
<td>Instruction materials</td>
<td>Checks for presence of lesson plan or quality of lesson plan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>2.23</td>
<td>Students' work (e.g. books/written work)</td>
<td>Checks for presence of instruction materials or quality of lesson plan</td>
<td></td>
</tr>
<tr>
<td>3.10</td>
<td>Graded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.11</td>
<td>Final grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.12</td>
<td>Catergoric grade</td>
<td>Gives final/overall grade to teacher</td>
<td></td>
</tr>
<tr>
<td>3.20</td>
<td>Discussion with Students</td>
<td>Gives overall grades by indicator category/time</td>
<td></td>
</tr>
<tr>
<td>4.10</td>
<td>Checks classroom environment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.11</td>
<td>Tables and chairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.12</td>
<td>Wall decorations</td>
<td>Checks for number of/quality of tables and chairs</td>
<td></td>
</tr>
<tr>
<td>4.13</td>
<td>Instruction materials e.g. chalkboard/chalk</td>
<td>Checks for number of/quality of/specific type of wall decorations</td>
<td></td>
</tr>
<tr>
<td>4.14</td>
<td>Children’s materials e.g. books/pens</td>
<td>Checks for number of/quality of/specific type of teaching materials</td>
<td></td>
</tr>
<tr>
<td>4.15</td>
<td>Classroom setup</td>
<td>Checks for number of/quality of/specific type of children's materials</td>
<td></td>
</tr>
<tr>
<td><strong>CC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.10</td>
<td>Classroom management</td>
<td>Degree to which the teacher arranges the classroom and learning materials in a way that is conducive to open discussion, is clean, and fosters a safe and welcoming learning environment.</td>
<td></td>
</tr>
<tr>
<td>5.11</td>
<td>Behaviour management + discipline</td>
<td>Degree to which the teacher states rules and expectations, monitors student behavior, anticipates behavioral problems, or subtly redirects misbehavior.</td>
<td></td>
</tr>
<tr>
<td>5.12</td>
<td>Time management</td>
<td>Degree to which the teacher efficiently completes managerial tasks like taking attendance, minimizes time spent on disruptions, provides ongoing tasks for students, and uses time productively.</td>
<td></td>
</tr>
<tr>
<td>5.13</td>
<td>Established routines</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.10</td>
<td>Feedback</td>
<td>Degree to which teachers institutes and upholds systemic classroom routines so that the class day follows a predictable pattern from which the students can easily follow.</td>
<td></td>
</tr>
<tr>
<td>6.11</td>
<td>Feedback Loops</td>
<td>Information provided by the teacher regarding aspect of student’s performance or understanding (Hattie &amp; Timperley, 2007).</td>
<td></td>
</tr>
<tr>
<td>6.12</td>
<td>Specific Feedback Types</td>
<td>Degree to which the teacher and students engage in meaningful, back-and-forth feedback exchanges.</td>
<td></td>
</tr>
<tr>
<td>6.20</td>
<td>Checking for Student understanding</td>
<td>Teacher gives scaffolding or corrective feedback, or another type of specific feedback</td>
<td></td>
</tr>
<tr>
<td>6.21</td>
<td>Formal Assessment</td>
<td>The degree to which the teacher creates opportunities to informally or verbally assess student understanding of content and monitor student learning. We do not include homework as a form of check for student understanding.</td>
<td></td>
</tr>
<tr>
<td>6.22</td>
<td>Objective-driven assessment</td>
<td>Degree to which the teacher creates opportunities to formally or informally check student understanding of content and monitor student learning.</td>
<td></td>
</tr>
<tr>
<td>6.30</td>
<td><strong>Content Understanding</strong></td>
<td>Degree to which formative and summative assessments are linked to unit objectives.</td>
<td></td>
</tr>
<tr>
<td>6.31</td>
<td>Instructional scaffolding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.32</td>
<td>Connections to the real world</td>
<td>Degree to which the teacher or other students offer hints or other assistance to a student who is unable to produce a correct answer.</td>
<td></td>
</tr>
<tr>
<td>6.33</td>
<td>Connections to student ideas and interests</td>
<td>Degree to which the teacher connects lesson content to the real world or to students’ everyday lives.</td>
<td></td>
</tr>
<tr>
<td>6.34</td>
<td>Connections to prior content knowledge</td>
<td>Degree to which the teacher incorporates students’ ideas and interests to inform class activities and assignments.</td>
<td></td>
</tr>
<tr>
<td>6.35</td>
<td>Use of multiple representations or examples</td>
<td>Degree to which the teacher incorporates lesson content that is connected to students’ prior knowledge of the subject or skills related to the concepts being taught.</td>
<td></td>
</tr>
<tr>
<td>6.36</td>
<td>Opportunities to practice applying concepts/procedures</td>
<td>Degree to which the teacher illustrates a concept, idea, or procedure by providing multiple examples and non-examples or by asking for more than one representation or perspective on the concept.</td>
<td></td>
</tr>
<tr>
<td>6.40</td>
<td><strong>Clear Presentation and Lesson Structure</strong></td>
<td>Degree to which the teacher provides students with the opportunity to practice applying concepts, procedures, or skills, either independently or under supervision.</td>
<td></td>
</tr>
<tr>
<td>6.41</td>
<td>Clear learning objectives</td>
<td>Degree to which the teacher presents information in a clear and organized manner, stays on topic, keeps the lesson moving, or does not allow students to spend excessive time on one task.</td>
<td></td>
</tr>
<tr>
<td>6.42</td>
<td>Clear lesson sequence (introduction and conclusion)</td>
<td>Degree to which the teacher clearly communicates learning objectives or re-orientates students to the objectives; students are aware of the lesson’s purpose.</td>
<td></td>
</tr>
<tr>
<td>6.50</td>
<td><strong>Language Development</strong></td>
<td>The degree to which each topic or learning exercise has a clear introduction and conclusion, and is sequenced logically.</td>
<td></td>
</tr>
<tr>
<td>6.51</td>
<td>Use of local language</td>
<td>Degree to which teacher uses language accurately, encourages students to use language correctly, and assists them in the process.</td>
<td></td>
</tr>
<tr>
<td>6.60</td>
<td><strong>Lesson facilitation and discourse</strong></td>
<td>Degree to which teacher uses local language to help students grasp complex concepts in English (or the primary language of instruction).</td>
<td></td>
</tr>
<tr>
<td>6.61</td>
<td>A variety of teacher-directed instructional strategies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>Category</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>6.62</td>
<td>Discussion facilitation</td>
<td>Degree to which the teacher uses a variety of teacher directed strategies or approaches and examples to actively engage all students in the lesson.</td>
<td></td>
</tr>
<tr>
<td>6.63</td>
<td>Grade appropriate lesson materials</td>
<td>Degree to which the teacher and students facilitate back and forth discussion.</td>
<td></td>
</tr>
<tr>
<td>6.64</td>
<td>Teacher flexibility/tailor to differential needs of students</td>
<td>Lesson materials and teacher’s language are appropriate for the grade level, and are consistent with students’ needs and abilities (from CLASS).</td>
<td></td>
</tr>
<tr>
<td>6.65</td>
<td>Encouraging peer-to-peer interaction, or group work</td>
<td>Degree to which the teacher tailors or individualizes support to students or demonstrates student needs flexibility with the lesson to accommodate students’ ability levels or understanding.</td>
<td></td>
</tr>
<tr>
<td>7.10</td>
<td>Critical thinking</td>
<td>Degree to which teacher engages students to work together on the task at hand.</td>
<td></td>
</tr>
<tr>
<td>7.11</td>
<td>Thinking questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.12</td>
<td>Thinking tasks</td>
<td>Degree to which the teacher asks students questions that seek to explore, draw connections, or identify strategies for solutions - critical thinking.</td>
<td></td>
</tr>
<tr>
<td>7.13</td>
<td>Student thinking questions/answers</td>
<td>Degree to which the teacher sets students tasks that seek to explore, draw connections, or identify strategies for solutions - critical thinking.</td>
<td></td>
</tr>
<tr>
<td>7.20</td>
<td>Student focus</td>
<td>Degree to which students explain an approach to solving a problem, a response or answer, their thinking process, or the meaning of an answer - critical thinking.</td>
<td></td>
</tr>
<tr>
<td>7.21</td>
<td>Teacher listens actively and encourages ideas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.22</td>
<td>Student autonomy</td>
<td>Degree to which the teacher actively listens to student comments, responds appropriately, and encourages students’ ideas.</td>
<td></td>
</tr>
<tr>
<td>7.23</td>
<td>Student ease in educational environment</td>
<td>Degree to which students can exert choice, take leadership opportunities, and have responsibilities in the classroom.</td>
<td></td>
</tr>
<tr>
<td>7.24</td>
<td>Student engagement</td>
<td>Degree to which the teacher allows students to share their ideas in class, ask for help, or take risks.</td>
<td></td>
</tr>
<tr>
<td>8.10</td>
<td>Positive Social climate</td>
<td>The degree to which students are actively engaged in class activities most of the time, as demonstrated by students paying attention, raising hands, answering questions, participating in group work, or the like.</td>
<td></td>
</tr>
<tr>
<td>8.11</td>
<td>Teacher awareness and responsiveness</td>
<td>Degree to which the teacher demonstrates awareness of student needs or problems, anticipates issues that students might have, offers assistance or support to individual students, is effective in addressing student...</td>
<td></td>
</tr>
<tr>
<td>8.12</td>
<td>Student social competence</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SS**

**CC**
<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.13</td>
<td>Positive Teacher-student interaction</td>
<td>Degree to which students display self-regulation and interpersonal knowledge.</td>
</tr>
<tr>
<td>8.20</td>
<td>Equality and Inclusiveness</td>
<td>Degree to which the teacher and students are physically close to each other, engage in social and rapport conversation, display warm/supportive interactions, offer each other praise and encouragement, listen to each other, cooperate, use names and respectful language/tone, and smile or laugh. Degree to which the classroom is free of anger, harsh voices, physical aggression, teacher threats or physical control of students, disrespect, bullying, teasing, and sarcasm.</td>
</tr>
<tr>
<td>8.21</td>
<td>Gender inclusiveness</td>
<td>Degree to which a teacher purposefully treats all students in the same manner and absolves from favoring some students over others.</td>
</tr>
<tr>
<td>8.22</td>
<td>Racial inclusiveness</td>
<td>Degree to which the teacher fosters a classroom environment that is unbiased toward gender.</td>
</tr>
<tr>
<td>8.23</td>
<td>Disability inclusiveness</td>
<td>Degree to which the teacher fosters a classroom environment that is unbiased toward race.</td>
</tr>
<tr>
<td>8.30</td>
<td>Motivating the classroom</td>
<td>Degree to which the teacher fosters a classroom environment that is unbiased toward disabled students, or students with learning difficulties.</td>
</tr>
<tr>
<td>8.31</td>
<td>Communicating high expectations</td>
<td>Degree to which the teacher imparts high expectations for student learning or demonstrates high-quality work.</td>
</tr>
<tr>
<td>8.32</td>
<td>Recognition of effort</td>
<td>Degree to which teacher recognizes a student's efforts to help others, assist in the classroom, and go above and beyond in their schoolwork.</td>
</tr>
<tr>
<td>8.33</td>
<td>Student goal setting</td>
<td>Teacher encourages goal setting</td>
</tr>
<tr>
<td>8.34</td>
<td>Positive language and morale</td>
<td>The teacher uses positive language with students, fosters high morale</td>
</tr>
<tr>
<td>8.40</td>
<td>Family engagement</td>
<td>Teacher involves the students' family in the learning process, when appropriate.</td>
</tr>
<tr>
<td>9.10</td>
<td>Professionalism and respect</td>
<td>Note: This section includes recommendations for teacher.</td>
</tr>
<tr>
<td>10.1</td>
<td>Notes taken after observation</td>
<td>Includes goal setting for teacher or improvement plan.</td>
</tr>
<tr>
<td>10.3</td>
<td>Teacher self-evaluation</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B
Coding Process and Decisions

We are looking for the existence of the following indicators and sub-indicators in the classroom observation tools in our sample. We quote the sections of the tool which we use to justify our coding, and, where possible, we state the category or indicator number in the tool, from which we quote.

We code the indicators as one if they are low-inference, two if they are high-inference, and zero if we cannot find evidence that they exist in the tool. The same tool may have a mix of high and low inference indicators.

<table>
<thead>
<tr>
<th>Indicator Type: (sub-)Indicator number – (sub-)indicator name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
</tr>
<tr>
<td>Assigned Code</td>
</tr>
<tr>
<td>Country (tool ID)</td>
</tr>
<tr>
<td>Indicator/category number: “quote from tool – range of responses e.g. yes, no”. Additional comments.</td>
</tr>
</tbody>
</table>

High and Low Inference Indicators
We consider indicators low-inference if they exist in the tool, but are recorded using check boxes, or yes or no questions. We also consider indicators that are recorded as ratings, but do not have a rubric nor examples for rating decisions, as low-inference. We consider indicators high-inference if they exist in the tool, and are recorded using a ratings scale, for which the observer has to refer to a rubric, or example behaviours.

Stand-Alone Indicators
Some indicators are stand-alone, meaning they have definitions, and can be coded independently of their sub-indicators. Therefore, it is sometimes possible to receive a score greater than zero for an indicator without receiving a score greater than zero for any of the sub-indicators. However, the converse is never true. Nevertheless, half of the indicators are not stand-alone, and hence are only coded as non-zero, if at least one sub-indicator is also coded as non-zero. Generally, if any sub-indicator is coded as non-zero, it’s parent indicator is also coded with an at least equal, non-zero, value.

For example, feedback is a stand-alone indicator. We define feedback as information provided by the teacher regarding aspect of student’s performance or understanding (Hattie & Timperley, 2007). As figure A.1 shows, if tool A contains specific feedback, such as scaffolding feedback, it meets the definition provided above, and therefore is also coded as 1 for the overall indicator, feedback. However, a tool may contain neither specific feedback nor feedback loops, and still meet the definition above. As shown by tool B, that would also be coded as 1 under the overall feedback indicator. In the case of an indicator that is not stand-alone, and does not have an independent definition, coding occurs like Tool A in the example above, but never like tool B.
Indicators not Included in this appendix
Indicators and sub-indicators 0.10 through 4.15, and 10.10 through 10.30 are not detailed below, as these indicators are either purely descriptive, or used for data collection. For example, indicator 2.10 measures what, if any, data is collected during the observation, and sub-indicator 2.12 specifies whether the tool records student absences. Such indicators are only coded from zero to one, as the designation high-inference is not applicable.

Indicator 5.10 – Classroom management
This is not a stand-alone indicator, meaning we only code this as one if one of the sub-indicators is also coded as a one.

Coded as 1
Brunei
This tool received a one under sub-indicator 5.12.

The Gambia
This tool received a one under sub-indicator 5.11.

Kenya (18)
This tool received a one under sub-indicator 5.12.

Kenya (19)
This tool received a one under sub-indicator 5.12.

Kenya (20)
This tool received a one under both sub-indicator 5.11 and sub-indicator 5.12.

Lebanon
This tool received a one under both sub-indicator 5.11 and sub-indicator 5.12.

Liberia
This tool received a one under sub-indicator 5.11.

Malawi
This tool received a one under sub-indicator 5.12.

Pakistan (57)
This tool received a one under both sub-indicator 5.11 and sub-indicator 5.13.

Pakistan (58)
This tool received a one under sub-indicator 5.11.

Tajikistan (30)
This tool received a one under sub-indicator 5.12.

Uganda (27)
This tool received a one under both sub-indicator 5.11 and sub-indicator 5.12.

Uganda (28)
This tool received a one under sub-indicator 5.12.

SDI
This tool received a one under sub-indicator 5.11.

Coded as 2
Guyana
This tool received a two under both sub-indicator 5.11 and sub-indicator 5.12.

Philippines
This tool received a two under sub-indicator 5.11.

Tajikistan (29)
This tool received twos for all sub-indicators.

Vietnam
This tool received a two under sub-indicator 5.12 and sub-indicator 5.13.

CLASS
This tool received twos under all sub-indicators.

FTT
This tool received twos under all sub-indicators.

MQI
This tool received a two under both sub-indicator 5.11 and 5.12.

SCOPE
This tool received a two under all sub-indicators.

TIPPS
This tool received a two under sub-indicator 5.11.

UTOP
This tool received a two under both sub-indicator 5.11 and 5.12.
Coded as 0
We cannot find evidence of classroom management in the following tools:
Papua New Guinea (10), Papua New Guinea (11), Uganda (26).

Sub-indicator 5.11 – Behaviour management and discipline
Definition: The degree to which the teacher states rules and expectations, monitors student behaviour, anticipates behavioural problems, or subtly redirects misbehaviour.

Coded as 1
The Gambia
In category D, “classroom management”: “Established classroom rules and procedures”, “minimised distractions and responded well to disruptive behaviour”, “did not use corporal punishment or other forms of humiliating treatment”, “recognised and rewarded good behaviour”. Rated on a scale from 1 to 5, from “unsatisfactory” to “excellent”, however with no rubric.

Kenya (20)
“Teacher deals promptly and effectively with undesirable behaviour – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon
Under category “classroom management”, indicators 1 and 8: “Consistent and proactive discipline – remarks”, and “effective mediation between different opinions or conflicts – remarks”.

Liberia
Indicator 5.10: “Is the class well behaved? – yes, no”

Pakistan (57)
Indicator 9: “Classroom management”. To score a three, the teacher “handles any poor student behaviour issues without disrupting lesson”. Although rated from 0 to 3, the rating is done progressively as a checklist. There is no rubric. Therefore, we code this as one.

Pakistan (58)
Indicator 4: “Discipline – acceptable or competent, needs improvement, unacceptable”.

Uganda (27)
Category 3, indicator 8: “Quality of learner behaviour”. Rated on a scale from 1 to 4, but with no rubric or examples.

SDI
Under category “teacher demeanour”, indicator 32: “Teacher hitting, pinching, or slapping a child? – yes, no”.

Stallings

Indicator 6: “Discipline”. Coded as a yes, no question, with the observer indicating whether the whole class, a large group of students, a small group of students, or one student, are being disciplined.

Coded as 2

Guyana

Under the category “classroom and student management”: “Rules of classroom interaction are evident and encouraging to all learners”, and “learner behaviour is monitored and managed to support learning”. Both indicators are rated on a scale from 0 to 4 using a rubric.

Philippines

Indicator 12: “Manages student behaviour constructively by applying positive and non-violent discipline to ensure learning-focused environments”. Rated on a scale from 1 to 9 using a rubric.

Tajikistan (29)

Under category “behaviour management”, indicators II and III: “Rules and expectations”, “Monitoring behaviour”. Both indicators are rated from 1 to 7 using a rubric.

CLASS

Under category “classroom organization”, indicator 1: “Behaviour management”. Rated on a scale of 1 to 7 using a rubric.

FFT

Under category “classroom environment”, indicators 2a and 2d: “Creating an environment of respect and rapport”, and “managing student behaviour”. Both indicators are rated on a scale from 1 to 4 using a rubric. The former indicator includes measures of how well the teacher deals with “disrespectful behaviour” at each rating level.

MQI

Under category “whole lesson codes”, indicator 1: “Lesson time is used efficiently”. Rated on a scale from 1 to 5 using a rubric. This indicator includes the degree to which “behaviour management takes away from instructional time”, in each rating level.

SCOPE

Under category “classroom culture”, indicator 1: “Supportive learning environment”. Rated from 1 to 5 using a rubric. This indicator includes a measure of “the teacher’s strategies to resolve conflicts or manage noncompliant behaviour”.

TIPPS

Indicator 15A: “Teacher uses tone of voice to control students”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 15B, which is also rated on the same scale. This indicator and its converse measure whether the teacher “resorts to yelling or rebuking students who misbehave”.

UTOP

Under category “classroom environment”, indicators 1.3 and 1.4: “Classroom on-task”, and “classroom management”. Both indicators are rated on a scale from 1 to 5, using a rubric. Indicator 1.3 contains measures of instances of “off-track behaviour”, from “consistent off-task behaviour” at a rating of 1, to “no instances of off-task behaviour” at a rating of 5. In a similar manner, indicator 1.4 contains measures of “disruptive behaviour”, although not in all five points of the rating scale. Additionally, under category “implementation”, indicator 3.6: “Implementation safety”. Rated on a scale from 1 to 5, using a rubric. This indicator includes measures of “inappropriate behaviour” at all five points of the rating scale.

Coded as 0
We cannot find evidence of behaviour management and discipline in the following tools: Brunei, Kenya (18), Kenya (19), Malawi, Papua New Guinea (10), Papua New Guinea (11), Tajikistan (30), Uganda (26), Uganda (28), Vietnam.

Sub-indicator 5.12 – Time management
Definition: The degree to which the teacher efficiently completes managerial tasks like taking attendance, minimizes time spent on disruptions, provides ongoing tasks for students, and uses time productively.

Coded as 1
Brunei
Under category “base learning on well-designed content and methods”, indicator 2.4: “time resource management”.

Kenya (18)
“Did the teacher teach with a perky pace? – yes, no”

Kenya (19)
“Did the teacher teach with a perky pace? – yes, no”

Kenya (20)
Indicator 2.2: “lesson timeline”. Rated on a scale of 4, from “unsatisfactory, satisfactory, good, very good”. Additionally, “teacher effectively manages time – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon

Malawi
Indicator 11: “teaching for effective learning”. To score a 2 or above on this indicator, teachers need to “use lesson time effectively to help children learn”. Although this indicator is rated from 1 to 4, it is done by a series of yes-no questions, therefore we code it as one.

Tajikistan (30)
Indicator 22: “teacher gives students enough time to practice learning new skills – yes, no, difficult to determine”

Uganda (27)
Category 3, indicator 7: “The pace of learning and use of time”. Rated on a scale from 1 to 4, but with no rubric or examples.

Uganda (28)
Under “lesson structure and content”, indicator 4: “Lesson started on time”. Rated from 1 to 4 but with no rubric or examples.

Coded as 2
Guyana
Under the category “classroom and student management”: “Lesson is paced for completion in time allocated”. Rated on a scale from 0 to 4 using a rubric.

Tajikistan (29)
Under category “lesson strategies and delivery”, indicator VII: “pace and time”. Rated from 1 to 7 using a rubric.

Vietnam
Indicator 6: “Time distribution and transitions”. Rated on a scale of 1 to 3, using a rubric.

CLASS
Under category “classroom organisation”, indicator 2: “Productivity”. Rated on a scale of 1 to 7 using a rubric. In this indicator, productivity “considers how well the teacher manages instructional time and routines so that the students have the opportunity to learn”.

FFT
Under category “planning and preparation”, indicator 1e: “Designing coherent instruction”. Rated on a scale from 1 to 4 using a rubric. To increase ratings from 1 to 3, teachers have to progress from “unreasonable time allocations”, to “reasonable time allocations”. Additionally, under category “the classroom environment”, indicator 2e: “Managing classroom procedures”. Also rated on a scale from 1 to 4 using a rubric. Each rating includes a measure of loss of instructional time, ranging from “much instructional time is lost due to inefficient classroom routines and procedures” at a rating of 1, to “instructional time is maximised due to efficient and seamless classroom routines and procedures” at a rating of 4.

MQI
Under category “whole lesson codes”, indicator 1: “Lesson time is used efficiently”. Rated on a scale from 1 to 5 using a rubric.

SCOPE
Under section I, “classroom structure”, indicators 3 and 6: “Participation of all learners”, and “manages reading and writing instruction”. Both indicators are rated on a scale from 1 to 5 using a rubric. Each rating level of indicator 3 contains measures of the amount of “wait time” during the lesson. Similarly, each rating level of indicator 6 includes measures of the amount of “non-productive use of learner time or disruption”.

UTOP
Under category “implementation”, indicator 3.4: “Implementation timing”. Rated on a scale from 1 to 5 using a rubric.

Coded as 0
We cannot find evidence of time management in the following tools:
The Gambia, Liberia, Papua New Guinea (10), Pakistan (57), Pakistan (58), Papua New Guinea (11), Philippines, Uganda (26), SCOPE, SDI, Stallings, TIPPS.

Sub-indicator 5.13 – Established routines
Definition: The degree to which the teacher institutes and upholds systemic classroom routines so class time follows a predictable pattern from which the students can easily follow.

Coded as 1
Pakistan (57)
Indicator 9: “Classroom management”. To score above a two, the teacher “has established norms and routines for students regarding classroom conduct”. Although rated from 0 to 3, the rating is done progressively as a checklist. Therefore, we code this as one.

Coded as 2
Tajikistan (29)
Under category “lesson strategies and delivery”, indicator VII: “pace and time”. Rated from 1 to 7 using a rubric. To score highly on this indicator, teachers need to have established routines – “everybody knows what is expected of them and how to go about doing it”

Vietnam
Indicator 13: “Routines”. Rated on a scale of 1 to 3, using a rubric.

CLASS
Under category “classroom organisation”, indicator 2: “Productivity”. Rated on a scale of 1 to 7 using a rubric. In this indicator productivity “considers how well the teacher manages instructional time and routines so that the students have the opportunity to learn”.
FFT
Under category “the classroom environment”, indicator 2c: “Managing classroom procedures”. Rated on a scale from 1 to 4 using a rubric. Each rating level includes a measure of routines, ranging from “much instructional time is lost due to inefficient classroom routines and procedures” at a rating of 1, to “instructional time is maximised due to efficient and seamless classroom routines and procedures” at a rating of 4.

SCOPE
Under section I, “classroom structure”, indicator 1: “Supportive learning environment”. Rated on a scale from 1 to 5 using a rubric. Each rating level of this indicator includes the extent to which “rules and routines” exist and are followed.

Coded as 0
We cannot find evidence of established routines in the following tools: Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Malawi, Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), MQI, SCOPE, SDI, Stallings, TIPPS, UTOP.

Indicator: 6.10 – Feedback
Definition: Information provided by the teacher regarding aspect of student’s performance or understanding (Hattie & Timperley, 2007).

Coded as 1
Brunei
Under category “use assessment and reporting effectively”, indicators 4.3a and 4.4: “Provide written and oral feedback”, and “communicate students’ achievements”.

The Gambia
In section E, “student engagement and learning”: “Teacher gave homework/assessed students’ learning, and provided feedback”. Rated from 1 to 5, however with no rubric.

Kenya (18)
“Did the teacher give immediate formative feedback throughout? – yes, sometimes, no”.

Kenya (19)
“Did the teacher give immediate formative feedback throughout? – yes, sometimes, no”.

Kenya (20)
“Teacher evaluates student’s answers – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”. “Teacher probes student’s answers – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”. “Teacher comments on student’s answers – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

“Teacher builds student answers into subsequent questions – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”. “Teacher moves around to interact with students to provide spoken and/or written feedback to inform learning – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon
Under the category “Classroom management”, indicator 5: “positive reinforcement and feedback – remarks”.

Pakistan (57)
Indicator 2: “assigning and checking written work”. To achieve a top score, teacher’s must provide “proper written feedback”. Indicator 6: “student assessment (oral/written)”. To achieve the top score, teachers must “offer solutions for incorrect answers and positive praise for correct answers”. Although these indicators are rated from 0 to 3, the rating is done progressively as a checklist. There is also no rubric. Hence, we code this as one.

Pakistan (58)
Under the category “The Pupils”: “Were the students given a chance to confirm correct answer? – yes, no, to some extent”. As this question follows directly from “were the students encouraged to ask questions – yes, no, to some extent”, it implies the giving of basic feedback.

Papua New Guinea (10)
Indicator 9h: “written corrections and feedback given by teacher – yes, no”.

Papua New Guinea (11)
Indicator 9h: “written corrections and feedback given by teacher – yes, no”.

Tajikistan (30)
Indicator 21: “teacher helps students to accomplish their learning tasks by: a) oral feedback b) written feedback…”. Indicator 25: “students’ activities and their learning outcomes are assessed through (mark all that apply): a) oral teacher’s feedback b) written teacher’s feedback…”.

Uganda (26)
Indicator 5: “Facilitation – giving feedback to participants”. Indicator 6: “Ability to assess TOTs effectively – timely feedback”. Both graded on a scale of three, from “needs development” to “well developed”, but without a rubric.

Uganda (27)
In category 4, “Assessment and record keeping”, indicator 3: “Clarity of feedback to learners indicating progress in achievement”. This is rated on a scale of four, from “poor” to “very good”, but with no rubric.

Uganda (28)
Indicator 16, under “Practical component of the lesson”: “Teacher gave feedback on learners’ exercises, assignments, and homework from previous lessons”. This is rated on a scale of four, from “weak” to “excellent”, but with no rubric.

Vietnam
Indicator 7: “Evaluation and feedback”. Rated on a scale of 1 to 3, using a rubric.

FFT
Indicator 3d: “Using assessment in instruction”. Rated on a scale from 1 to 4 singing a rubric. To score highly on this indicator, there needs to be “a variety of forms of feedback, from both teachers and peers, is accurate and specific, and advances learning”. As feedback is embedded in an assessment indicator, and does not have its own indicator, we code this as one.

SCOPE
Indicator 13: “Writing instruction”. Rated on a scale from 1 to 5, using a rubric. To score highly, teachers must “discuss work with learners and provide thoughtful feedback”. As feedback does not have its own indicator, and is embedded in a writing indicator, we code this as one.

SDI
This tool received a one under sub-indicator 6.12. Additionally, under category “feedback”, indicator 37: “the teacher gave feedback of praise, moral strengthening and/or encouragement – never, once, more than once”.

Coded as 2
Guyana
Under the category “Pedagogy”: “Learner ideas, experience, and feedback are solicited/integrated into the lesson”, and “Feedback provided to learners is useful, respectful, and timely”. Both indicators are rated from 0 to 4, using a rubric.

Philippines
Indicator 6: “Uses strategies for providing timely, accurate, and constructive feedback to improve learning performance”. Rated from 1 to 9, using a rubric.

Tajikistan (29)
This tool received a two for sub-indicator 6.12. Furthermore, under “assessment and evaluation”: indicators II and III, “Feedback approaches”, and “encouragement and affirmation”. Under III, the teacher must “consistently go beyond simply saying that a response is correct or incorrect”, and “give specific feedback that is individualised to specific students or contexts of learning”. Rated from 1 to 7.

CLASS
This tool received twos under both sub-indicator 6.11 and sub-indicator 6.12. Additionally, under the category “instructional support”: “Quality of feedback”, rated 1 to 7 using a rubric.

MQI
This tool received a two under sub-indicator 6.12.

**TIPPS**

This tool received a two under sub-indicator 6.12.

**Coded as 0**

We cannot find evidence of any feedback in the following tools:
Liberia, Malawi, SCOPE, Stallings, UTOP.

**Sub-indicator 6.11 – Feedback Loops**

Definition: Back-and-forth exchange, in which the teacher and students engage in meaningful feedback interactions.

**Coded as 1**

**Philippines**

To achieve a top score on indicator 6 of their tool, teachers need to “foster constructive feedback cycle[s] (teacher-students and students-students)” and have a “positive stance towards giving and receiving feedback”. We code this as one because it is embedded in a generic feedback indicator, among many other things also embedded in this indicator, and hence is only measured for top performing teachers.

**Coded as 2**

**Guyana**

Under the category “Pedagogy”: “Learner ideas, experience, and feedback are solicited/integrated into the lesson”, and “Feedback provided to learners is useful, respectful, and timely”. Both indicators are rated from 0 to 4, using a rubric, and taken together measure both feedback from the teacher to the students, and the students to the teacher.

**CLASS**

Under the category “instructional support”: “Quality of feedback”. Rated 1 to 7 using a rubric. This indicator includes a measure of “feedback loops”.

**Coded as 0**

We cannot find evidence of feedback loops in the following tools:
Brunei, The Gambia, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Papua New Guinea (10), Papua New Guinea (11), Pakistan (57), Pakistan (58), Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, FFT, MQI, SCOPE, Stallings, UTOP.

**Sub-indicator 6.12 – Specific Feedback Types**

Here we look for differentiation between generic feedback, as in 6.10, and different types of feedback such as corrective or scaffolding feedback.

Definition: Scaffolding feedback is feedback that expands on or clarifies a student’s answer, and corrective feedback as feedback intended to correct or identify an error in the student’s answer. However, we do not include yes or no responses, even if corrective or scaffolding, and code those under indicator 6.10.
Kenya (20)
“Teacher builds student answers into subsequent questions – yes, no”. This indicator expands on feedback given to students, and builds it into the lesson, which is an element of scaffolding feedback.

Vietnam
Indicator 5: “class activity”. To achieve a top score, “class activity [must be] guided by the teacher but conducted by a student or a group of students – teacher observes and provides feedback and input as needed”. The latter implies aspects of scaffolding feedback, and feedback that goes beyond traditional teacher-student exchanges.

SDI
Under category “feedback”, indicator 38: “teacher gave feedback that was correcting a mistake – never, once, more than once”. This is corrective feedback.

Tajikistan (29)
Under “assessment and evaluation”, indicator II: “feedback approaches”. To achieve a top score on this indicator, teachers must “usually provide scaffolding for students who are having a hard time”. The feedback indicator rubric contrasts scaffolding feedback to other feedback, and only the degrees of timeliness and scaffolding feedback to score the teacher. Hence, we code this as two. Rated from 1 to 7.

CLASS
Under the category “instructional support”: “Quality of feedback”. Rated 1 to 7 using a rubric. This indicator includes a measure of “scaffolding”.

MQI
Under category “working with students and mathematics”, Indicator 1: “remediation of student errors and difficulties”. It is rated on a scale with a range of four, from “not present” to “high”, using a rubric. They define this indicator further as “procedural remediation corrects student problems with procedures...To score an instance of procedural remediation, there must be more than a simple correction of a student mistake”. This contains elements of corrective feedback.

TIPPS
Indicator 7B: “Teacher provides pupils with specific feedback to facilitate learning rather than just getting the correct answer or finishing an activity”. Indicator 10B: “Teacher repeats and/or extends pupil responses to promote deeper understanding and learning of a concept”. These both contain elements of scaffolding feedback. Both indicators are rated using a rubric, from “somewhat accurate” to “very accurate”, against their converse, indicators 7A and 10A respectively, which are also rated on the same scale.

We cannot find evidence of specific feedback types in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Lebanon, Liberia, Malawi, Papua New Guinea (10), Papua New Guinea (11), Pakistan (57), Pakistan (58), Philippines, Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), FFT, SCOPE, Stallings, UTOP.

**Indicator 6.20 – Checking for student understanding**
Definition: The degree to which the teacher creates opportunities to informally or verbally assess student understanding of content and monitor student learning. We do not include homework as a form of check for student understanding.

**Coded as 1**

**Brunei**
This tool received a one for sub-indicator 6.21.

**The Gambia**
Under section D, “classroom management”: “Moved around the classroom to check pupils’ work”. Under section E, “student engagement and learning”: “teacher asked questions to promote students’ critical thinking and check understanding”, and “teacher gave homework/assessed students’ learning”. These indicators are rated from 1 to 5, however with no rubric.

**Kenya (18)**
“What comprehension (after reading the story) activities were modelled? [tick all that apply] – checked prediction, asked questions from the script, asked pupils to retell the story, other (specify in the next section), not applicable, skipped this section”. “Did the teacher ask questions from the story? – yes, no”. “Did the teacher make the children try to go back to the test to find the answers to the questions? – yes, not applicable, no”.

**Kenya (20)**
“Teacher asks closed questions requiring students to recall repeat information – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”, “teacher asks open ended questions – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”, “teacher calls on students to answer questions individually – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”, “teacher asks students to demonstrate in front of the class – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

**Lebanon**
This tool received a one under sub-indicator 6.21.

**Liberia**
This tool received a one for sub-indicators 6.21 and 6.22.

**Malawi**
Indicator 11: “Teaching for effective learning”. To score a level three of above on this indicator, 75% of teacher's must: “use questioning appropriately to check students’ understanding”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, therefore we code it as one.

Pakistan (57)
This tool received a one under sub-indicator 6.21. Additionally, on indicator 6: “student assessment (oral/written). To score at least a one on indicator 6, teachers must “ask relevant oral questions”. Although these indicators are rated from 1 to 3, the rating is done progressively as a checklist. There is also no rubric. Hence, we code this as one.

Pakistan (58)
Under the category “the pupils”: “did the teacher ask focused questions? – yes, no, to some extent”

Papua New Guinea (10)
Indicator 6e: “teacher asking comprehension questions when reading story/text – yes, no”.

Papua New Guinea (11)
Indicator 6e: “teacher asking comprehension questions when reading story/text – yes, no”.

Tajikistan (30)
This tool received a one under sub-indicator 6.21.

Uganda (26)
This tool received a one for sub-indicator 6.21.

Uganda (27)
This tool received a one in sub-indicators 6.21 and 6.22. Additionally, in category 3, “Quality of teaching and learning process”, indicator 6: “The clarity and purposefulness of questioning”. This is rated on a scale of four, from “poor” to “very good”, but with no rubric.

Uganda (28)
Indicators 8 and 9 under “teacher-student interactions and participation”: “teacher checked students’ understanding of the lesson using questions and answers”, and “teacher encouraged students to ask him/her questions”. This is rated on a scale of four, from “poor” to “very good”, but with no rubric.

SDI
Under category “teacher asking questions”, indicators 33 and 34. “The teacher asked questions that required learners to recall information – yes, no”. “The teacher asked learners to carry out a task which allowed them to demonstrate their understanding of what they had learned during the lesson – yes, no”.

Guyana
This tool received twos under sub-indicators 6.21 and 6.22. Under the category “Pedagogy”: “Appropriate questioning techniques are used”. Rated from 0 to 4, using a rubric.

Philippines
This tool received a two for sub-indicator 6.21

Tajikistan (29)
This tool received a two for sub-indicator 6.21. Additionally, to score highly under “Assessment and evaluation”, indicator I: “Assessment Strategies”, teachers need to “consistently check for student understanding and support students in need”, and “consistently perform a pre-assessment of students’ skills, knowledge, or experience”. Rated from 1 to 7.

Vietnam
Indicator 6: “evaluation and feedback”. To score highly on this indicator, teachers need to make sure “multiple forms of evaluation are used throughout the lesson – from self-evaluation to teacher’s evaluation – reflection takes place on evaluation process”. Indicator 8: “Use of inquiry mode”. To score highly on this indicator, teachers need to make sure “enquiry is the default mode of teaching and learning, students and teacher engage in respectful questions and counter-questions to develop understanding”. Both these indicators are rated on a scale from 1 to 3, using a rubric.

FFT
This tool received twos under both sub-indicator 6.21 and sub-indicator 6.22.

SCOPE
Under Section I, “classroom culture”, indicator 4: “opportunities for reflection”. Rated on a scale from 1 to 5 using a rubric. To score highly, “learners are provided regular, extended opportunities to assess their own learning and to set improvement goals and monitoring strategies”.

UTOP
This tool received twos under both sub-indicator 6.21 and sub-indicator 6.22. Additionally, category 2, “lesson structure”, indicator 2.3: “lesson assessments: the structure of the lesson included opportunities for the instructor to check student understanding”. Category 3, “implementation”, indicator 3.1: “implementation questioning: the teacher used questioning strategies to encourage participation, check on skill development”. Both indicators are rated on a scale of 1 to 5 using a rubric.

We cannot find evidence of any checks for student understanding in the following tools: Kenya (19), CLASS, MQI, Stallings, TIPPS.
Sub-indicator 6.21 – Formal Assessment
Definition: The degree to which the teacher creates opportunities to formally assess student understanding of content and monitor student learning, for example through assessments, tests, or quizzes.

Coded as 1
Brunei
In the category “professional knowledge and skills”, indicator 4: “Use assessment and reporting effectively – + / –”.

Kenya (20)
Indicator 1.5: “Assessment of learning – unsatisfactory, satisfactory, good, very good”.

Lebanon
Under the category “lesson plan” indicator 5: “Identification of the assessment methods”. Under the category “methods of assessment”, indicators 1 to 7: “questions and answers”, “class discussion”, “group response”, “project/product”, “writing product”, “review of main concepts”, “quiz/test”. It is unclear how these are recorded. They appear to be a checklist, with space for remarks at each category.

Liberia
Indicators 5.5: “was the teacher able to show you up-to-date assessment records for every student? – yes, no”. The keeping of records implies formal assessments.

Pakistan (57)
Indicator 6: “student assessment (oral/written). To score at least a one on indicator 6, teachers must “conduct relevant written assessment to test students’ understanding”. Although these indicators are rated from 1 to 3, the rating is done progressively as a checklist. There is also no rubric. Hence, we code this as one.

Tajikistan (30)
Indicator 25: “Students’ activities and their learning outcomes are assessed through (mark all that apply) – a) oral teacher’s feedback, b) written teacher’s feedback, c) peer-assessment d) self-assessment e) student reflection on lesson’s results f) given marks by the teacher g) not assessed, h) difficult to determine, i) other”. Indicator 23: “the lesson plan includes students’ assessment activity (method) – a) oral teacher’s feedback, b) written teacher’s feedback, c) peer assessment, d) self-assessment, e) student reflection on lessons results, f) given marks by the teacher, g) nothing, h) other”

Uganda (26)
In category 6, “Ability to assess TOTs effectively, indicators: “continuity in assessment”, and “Assess TOTs confidently”. Graded on a scale of three, from “needs development” to “well developed”, but without a rubric.

Uganda (27)
In category 4, “Assessment and record keeping”, indicators 1, 4, 5, and 7: “assessment successfully evaluates a range of competencies”, “assessment is integrated in learning
process”, “assessment can inform teachers’ planning”, “assessment records are safely kept and easy to retrieve”. Graded on a scale of three, from “needs development” to “well developed”, but without a rubric.

**Coded as 2**

**Guyana**

Under the category “Planning”: “Assessment strategies are identified and detailed in the plan”. Rated from 0 to 4, using a rubric.

**Philippines**

Indicator 5: “Designs, selects, organises, and uses diagnostic, formative, and summative assessment strategies consistent with curriculum requirements”. Rated from 1 to 9 using a rubric.

**Tajikistan (29)**

Under “Assessment and evaluation”, indicator I: “Assessment Strategies”. Rated from 1 to 7 using a rubric.

**FFT**

Indicator 1F: “Designing student assessments”. Rated on a scale from 1 to 4 using a rubric. To score highly, “assessment is fully integrated into instruction through extensive use of formative assessment”.

**UTOP**

Category 3, “implementation”, indicator 3.3: “implementation modification: the teacher used formative assessment effectively to be aware of the progress of all students”. Rated on a scale of 1 to 5 using a rubric.

**Coded as 0**

We cannot find evidence of any formal assessment in the following tools:

Kenya (18), Kenya (19), The Gambia, Malawi, Pakistan (58) Papua New Guinea (10), Papua New Guinea (11), Uganda (28), Vietnam, CLASS, MQI, SCOPE, Stallings, TIPPS.

**Sub-indicator 6.22 – Objective Driven Assessment**

Definition: The degree to which formative and summative assessments are linked to some standard, or to unit objectives.

**Coded as 1**

**Liberia**

Indicator 5.6: “are the methods of student assessment consistent with the recommended standards in the LEAR? – yes, no”. This indicator implies that assessments are linked to certain objectives or standards.

**Philippines**

To achieve a good score on indicator 5 of their tool, teachers need to “integrate assessment strategies consistent with the curriculum requirements”. We code this as one because it is embedded in a generic feedback indicator, among many other things.
also embedded in this indicator, and hence is only measured for mid to top performing teachers.

Tajikistan (30)
Indicator 27: “planned assessment tools are appropriate to measuring learning objectives – yes, no, partly”.

Uganda (27)
Category 4, “assessment and record keeping”, indicator 2: “assessment matches criteria set out in national curriculum guidelines”. Graded on a scale of three, from “needs development” to “well developed”, but without a rubric.

Coded as 2
Guyana
Under the category “Planning”: “Objectives, content and assessments are aligned”. Rated from 0 to 4, using a rubric.

FFT
Indicator 1F: “Designing student assessments”. Rated on a scale from 1 to 4 using a rubric. To score highly, “all the instructional outcomes may be assessed by the proposed assessment plan”.

UTOP
Category 4, “mathematics/science content”, indicator 4.4: “content assessments: formal assessments used by teacher (if available) were consistent with content objectives”. Rated on a scale of 1 to 5 using a rubric.

Coded as 0
We cannot find evidence of objective driven assessment in the following tools: Brunei, Kenya (18), Kenya (19), Kenya (20), The Gambia, Lebanon, Malawi, Papua New Guinea (10), Pakistan (57), Pakistan (58), Papua New Guinea (11), Tajikistan (29), Uganda (26), Uganda (28), Vietnam, CLASS, MQI, SCOPE, Stallings, TIPPS.

**Indicator 6.30 – Content understanding**
This is not a stand-alone indicator, meaning we only code this as one if one of the sub-indicators is also coded as a one.

Coded as 1
Brunei
This tool received a one under sub-indicator 6.36.

The Gambia
This tool received a one under sub-indicator 6.36.

Kenya (18)
This tool received a one under sub-indicator 6.36.
Kenya (19)
This tool received a one under sub-indicator 6.35.

Kenya (20)
This tool received a one under sub-indicator 6.31, sub-indicator 6.35, and sub-indicator 6.36.

Lebanon
This tool received a one for sub-indicator 6.32, sub-indicator 6.33, and sub-indicator 6.34.

Malawi
This tool received a one under sub-indicator 6.33, sub-indicator 6.34, and sub-indicator 6.36.

Pakistan (57)
This tool received a one under sub-indicator 6.34, sub-indicator 6.35, and sub-indicator 6.36.

Papua New Guinea (10)
This tool received a one for both sub-indicator 6.35 and sub-indicator 6.36.

Papua New Guinea (11)
This tool received a one for both sub-indicator 6.35 and sub-indicator 6.36.

Tajikistan (30)
This tool received a one for both sub-indicator 6.34 and sub-indicator 6.36.

Uganda (26)
This tool received a one for sub-indicator 6.35.

Uganda (27)
This tool received a one for sub-indicator 6.32, sub-indicator 6.34, sub-indicator 6.35, and sub-indicator 6.36.

Uganda (28)
This tool received a one for sub-indicator 6.32, sub-indicator 6.34, and sub-indicator 6.36.

SDI
This tool received a one for both sub-indicator 6.32 and sub-indicator 6.36.

Stallings
This tool received a one for sub-indicator 6.36.

Coded as 2
Guyana
This tool received a two under sub-indicator 6.36

Philippines
This tool received a two under both sub-indicator 6.34 and sub-indicator 6.35.

Tajikistan (29)
This tool received a two under sub-indicator 6.31, sub-indicator 6.32, and sub-indicator 6.34.

Vietnam
This tool received a two under sub-indicator 6.32, sub-indicator 6.33, and sub-indicator 6.34.

CLASS
This tool received twos for all sub-indicators.

FFT
This tool received a two under sub-indicator 6.31, sub-indicator 6.32, and sub-indicator 6.33.

MQI
This tool received a two under sub-indicator 6.33, sub-indicator 6.34, and sub-indicator 6.35.

SCOPE
This tool received a two under both sub-indicator 6.33 and sub-indicator 6.36.

TIPPS
This tool received a two under sub-indicator 6.33, sub-indicator 6.34, and sub-indicator 6.35.

UTOPI
This tool received a two under sub-indicator 6.31, sub-indicator 6.32, sub-indicator 6.33, sub-indicator 6.34, and sub-indicator 6.36.

Coded as 0
We cannot find evidence of content understanding in the following tools:
Liberia, Pakistan (58).

Sub-indicator 6.31—Instructional scaffolding
Definition: The degree to which the teacher or other students offer hints or other assistance to a student who is unable to produce a correct answer.

Coded as 1
Kenya (20)
“Teacher evaluates student answers – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”,
“teacher probes student answers – behaviour never observed, behaviour rarely
observed, behaviour occasionally observed, behaviour consistently observed”, “teacher comments on student answers – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”. These questions follow on from one another, and together have aspects of instructional scaffolding.

Coded as 2
Tajikistan (29)
Under category “assessment and evaluation”, indicator II: “Feedback approaches”. Rated on a scale from 1 to 7 using a rubric. This indicator includes measures of the degree to which the teacher “provides scaffolding to students”, at all rating levels.

CLASS
Under category “instructional support”, indicator 4: “Quality of feedback”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “providing hints”.

FFT
Under category “instruction”, indicator 3a: “Communicating with students”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “explanations of content”, which includes “appropriate scaffolding”.

TIPPS
Indicator 11B: “Teacher uses scaffolding to provide a step-by-step framework to help pupils learn and understand subject matter”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 11A, which is also rated on the same scale.

Coded as 0
We cannot find evidence of instructional scaffolding in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (30), Uganda (26), Uganda (27), Vietnam, MQI, SCOPE, UTOP, Stallings.

Sub-indicator 6.32 – Connections to the real world
Definition: The degree to which the teacher connects lesson content to the real world or to students’ everyday lives.

Coded as 1
Lebanon
Under category “building background”, indicator 1: “linking new concepts to students’ backgrounds and experiences – remarks”.

Philippines
Indicator 3: “Selects, develops, organises, and uses appropriate teaching and learning resources, including ICT, to address learning goals”. Indicator 4: “Ensures the positive use of ICT to facilitate the teaching and learning process”. Indicator 10: “Maintains
supportive learning environments that nurture and inspire learners to participate, cooperate, and collaborate in continued learning”. Indicator 11: “Applies a range of successful strategies that maintain learning environments that motivate learners to work productively by assuming responsibility for their own learning”. Indicator 21: “Plans, manages, and implements developmentally sequenced teaching and learning process to meet curriculum requirements and varied teaching contexts”. All these indicators are rated on a scale from 1 to 9 using a rubric. To score highly on all of the above indicators, teachers need to establish “real world connections”. Although these are rated on a scale using a rubric, for each indicator, establishing real world connections is only measured for high-scoring teachers. Therefore, we code this as a one.

Uganda (27)
Under category “teacher knowledge”, indicator 4: “ability of the teacher to contextualise knowledge with locally relevant examples”. Rated on a scale from 1 to 4 but with no rubric nor examples. This implies a connection to the students’ everyday lives and their world.

Uganda (28)
Under category “practical component of the lesson”, indicator 14: “Teacher used practical/real life examples”. Rated on a scale from 1 to 4 but with no rubric nor examples.

MQI
Under category “common core aligned student practices”, indicator 5: “Students work with contextualised problems”. Rated on a scale of 4, from “not present” to “high”. This indicator measures whether students are given “contextualised problems (e.g. story problems, real-world applications...”). However, the rubric measures the amount of scaffolding from the teacher rather than the degree of the real-world applicability of the tasks. Therefore, it acts as a yes or no check for contextualised problems, and we code this as one.

SDI
Under category “use of teaching aids during the lesson”, indicator 25: “the teacher used local information to make learning relevant – yes, no”.

Coded as 2
Tajikistan (29)
Under category “lesson strategies and delivery”, indicator V: “Effective use of relevance pedagogy (RP) strategies”. Rated on a scale from 1 to 7 using a rubric. This indicator measures the degree to which the teacher “relate[s] concepts to the students’ actual lives”.

Vietnam
Indicator 23: “Life Skills”. Rated on a scale of 3, from “basic” to “advanced”, using a rubric. This indicator measures the “connections of activities to the real world” and their “application”. Additionally, indicator 22: “Use of real artefacts”. Also rated on a scale of 3, from “basic” to “advanced”.
CLASS (K-3 only)
Under category “instructional support”, indicator 1: “Concept development”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “connections to the real world” at every rating level.

FFT
Under category “instruction”, indicator 3a: “Communicating with students”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of the degree to which teachers “connect…explanations to students’ interests and lives beyond school”.

TIPPS
Indicator 6B: “Teacher connects pupils’ studies to their everyday life experiences, showing the relevance of lessons outside the classroom”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 6A, which is also rated on the same scale.

UTOP
Under category “mathematics/science content”, indicator 4.8: “Content societal impact”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the degree to which “there was discussion about the content topic’s role in history, current events, or relevant “real-world” problems”.

Coded as 0
We cannot find evidence of connections to the real world in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Liberia, Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Uganda (26), Tajikistan (30), CLASS, SCOPE, Stallings.

Sub-indicator 6.33 – Connections to student ideas and interests
Definition: The degree to which the teacher incorporates students' ideas and interests to inform class activities and assignments.

Coded as 1
Lebanon
Under category “classroom management” indicator 2: “Regular incorporation of students’ ideas into the lesson – remarks”.

Malawi
Indicator 1: “learning in lessons”. To score a 3 or above on this indicator, students must be “interested and want to learn”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, so we code it as one.

Philippines
Indicator 13: “Uses differentiated, developmentally appropriate learning experiences to address learners’ gender, needs, strengths, interests, and experiences”. Rated on a
scale from 1 to 9 using a rubric. However, although the title of the indicator refers to students’ interests, the rubric has no mention of this, only of “differentiated and developmentally appropriate learning experiences”. Therefore, we code this as one.

Coded as 2
Vietnam

Indicator 12: “Flexibility and student focus”. Rated on a scale of 3, from “basic” to “advanced”, using a rubric. This indicator measures whether the teacher “incorporates student’s ideas” into instruction. Additionally, indicator 21: “Creativity”. This indicator measures the degree to which the teacher creates opportunities for students to “generate their own ideas and products”.

CLASS
Under category “emotional support”, indicator 3: “Regard for student perspectives”. Rated on a scale from 1 to 7 using a rubric. This indicator measures “the degree to which the teacher’s interactions with students and classroom activities place an emphasis on students’ interests, motivations, and points of view”.

FFT
Under category “instruction”, indicator 3a: “Communicating with students”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of the degree to which teachers “connect…explanations to students’ interests and lives beyond school”.

MQI
Under category “whole lesson codes”, indicator 6: “Teacher uses student ideas”. Rated on a scale from 1 to 5 using a rubric. This indicator “describes the extent to which the teacher uses student ideas and solutions to move the lesson forward”.

SCOPE
Under category “language and literacy instruction”, indicator 7: “Opportunities for oral language development”. Rated on a scale from 1 to 5 using a rubric. This indicator contains measures for how far students are encouraged to “express ideas and opinions” and the how far the teacher “facilitates the flow of discussion”. Additionally, under category “language and literacy instruction”, indicator 8: “Opportunities for meaningful reading”. Also rated on a scale from 1 to 5 using a rubric. This indicator measures the extent to which the teacher “provides a variety of texts and allows the learners to choose their own reading matter. Texts match learners’…interests”.

TIPPS
Indicator 2B: “Teacher uses pupils’ ideas and interests to inform class activities and assignments”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 2A, which is also rated on the same scale.

UTOP
Under category “classroom environment”, indicator 1.1: “Classroom engagement”. Rated on a scale from 1 to 5 using a rubric. Classroom engagement is defined as the degree to which “the classroom environment facilitated by the teacher encouraged
students to generate ideas, questions, conjectures, and/or propositions that reflected engagement or exploration”. The rubric for this indicator specifies teacher “facilitation” of this process, which implies that the teacher incorporates these ideas into the lesson.

Coded as 0
We cannot find evidence of connections to student ideas and interests in the following tools: Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Liberia, Papua New Guinea (10), Pakistan (58), Papua New Guinea (11), Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), SDI, Stallings.

Sub-indicator 6.34 – Connections to prior content knowledge
Definition: The degree to which the teacher incorporates lesson content that is connected to students’ prior knowledge of the subject or skills related to the concepts being taught.

Coded as 1
Kenya (20)
“Teacher checks for prior knowledge – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon
Under category “building background”, indicator 2: “linking new concepts to part learning skills – remarks”.

Malawi
Indicator 11: “teaching for effective learning”. To score a 4 on this indicator, teachers need to “explain to students what they are going to learn and its links with what they already know”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, so we code it as one.

Pakistan (57)
Under category “evaluation during lesson”, indicator 3: “SLO (Taleemi calendar)”. To score a three on this indicator, teachers need to “explain connections to the previous lesson”. Although rated from 0 to 3, the rating is done progressively as a checklist. There is no rubric. Therefore, we code this as one.

Tajikistan (30)
During the examination of the lesson plan, indicator 17: “Lesson plan describes students’ prior knowledge – yes, no”.

Uganda (27)
Under category “teacher knowledge”, indicator 3: “Ability to relate different areas of subject content in coherent way”. Rated on a scale from 1 to 4, but with no rubric nor examples. This implies the presence of some connection between subject areas and prior knowledge.

Uganda (28)
Under category “lesson structure and content”, indicator 4: “The objectives of the lesson and the relation with previous/next lessons were clearly explained”. Rated on a scale from 1 to 4 but with no rubric nor examples.

FFT
Under category “planning and preparation”, indicators 1c and 1e: “setting instructional outcomes”, and “designing coherent instruction”. Both indicators are rated on a scale from 1 to 4 using a rubric. To score highly on both indicators, teachers need to “connect outcomes to previous and future learning” (1c) or make “learning experiences connect to other disciplines” (1e). However, these indicators only measure connections to prior learning for high-scoring teachers, as such connections are not among the four key parts of either indicator. Therefore, we code this as one.

**Coded as 2**

Philippines
Indicator 1: “Applies knowledge of content within and across curriculum content teaching areas”. Rated on a scale from 1 to 9 using a rubric. The reference to “across curriculum content teaching areas” is sustained through the rating scale, and implies connections to students’ previous knowledge, and knowledge from other classes. Additionally, indicator 21: “Plans, manages, and implements developmentally sequenced teaching and learning process to meet curriculum requirements and varied teaching contexts”. To score highly on this indicator teachers need to manage “well-structured lessons with emphasis on explicit connections between previous learning and new concepts and skills”.

Tajikistan (29)
Under category “lesson strategies and delivery”, indicator 1: “Effective facilitation”. Rated on a scale from 1 to 7 using a rubric. This indicator includes the degree to which the teacher “links concepts and activities to one another and to previous learning”, at each rating level.

Vietnam
Indicator 24: “Cross-subject unification”. Rated on a scale of 3, from “basic” to “advanced”, using a rubric. This indicator measures the teacher’s “references to other subjects”, and the “clarity of the connections”.

CLASS (K-3 only)
Under category “instructional support”, indicator 1: “Concept development”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “integration with previous concept” at every rating level.

MQI
Under category “richness of the mathematics”, indicator 1: “Linking between representations”. Rated on a scale of 4 from “not present” to “high” using a rubric. This indicator measures “explicit linking and connections between different representations of a mathematical idea or procedure”.

SCOPE
Under category “classroom culture”, indicator 3: “Participation of all learners”. Rated on a scale from 1 to 5 using a rubric. This includes a measure of the degree to which the teacher “the teacher orchestrates the class such that the prior knowledge and personal interests are used as the basis for conversations, activities, and learning experiences”.

TIPPS

Indicator 5B: “Teacher connects activities and subject matter to achieve larger instructional concepts and objectives”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 5A, which is also rated on the same scale. The rubric for this indicator specifies connections to “prior/upcoming lessons” or “prior/upcoming concepts”.

UTOP

Under category “implementation”, indicator 3.5: “Implementation connections”. Rated on a scale from 1 to 5 using a rubric. Implementation connections are established if “instructional strategies and activities used in this lesson [are] clearly connected to students’ prior knowledge and experience”. Additionally, under category “mathematics/science content”, indicator 4.7: “Content interconnections”. Also rated on a scale from 1 to 5 using a rubric.

Coded as 0

We cannot find evidence of connections to prior content knowledge in the following tools: Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Liberia, Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Uganda (26), CLASS, SDI, Stallings.

Sub-indicator 6.35 – Use of multiple representations and examples

Definition: The degree to which the teacher illustrates a concept, idea, or procedure by providing multiple examples and non-examples or by asking for more than one representation or perspective on the concept.

Coded as 1

Kenya (19)

In multiple instances of the tool, questions such as “what ______ activities were taught? – list of activities”. For example, “what counting activities were taught? – forwards, backwards, skip counting (by 2s, 5s, 10,) start from a number that is not 1, others”. These are all different methods of practicing and applying counting knowledge, and the indicator looks for the presence of variety.

Kenya (20)

Under category “scheme of work”, indicator 1.3: “range of teaching and learning activities – unsatisfactory, satisfactory, good, very good”. Additionally, “Teacher uses a range of instruction materials – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Pakistan (57)
Under category “evaluate during the lesson”, indicator 5: “Visual aids (support material)”. To score a 3 on this indicator, the teacher needs to use “two or more visual aids which are instructive and relevant to SLOs along with using board”. Although rated from 0 to 3, the rating is done progressively as a checklist. There is no rubric. Therefore, we code this as one.

Papua New Guinea (10)
Under category “teaching”, indicator 6i: “Varied reading activities during lesson (e.g. spelling, verbal reporting, general discussions… – yes, no”.

Papua New Guinea (11)
Under category “teaching”, indicator 6i: “Varied reading activities during lesson (e.g. spelling, verbal reporting, general discussions… – yes, no”.

Uganda (26)
Under category “facilitation”, indicators 2 and 9: “use of examples”, and “ability to demonstrate practice activities”. Both graded on a scale of three, from “needs development” to “well developed”, but with no rubric nor examples.

Uganda (27)
Under category “quality of teaching and learning process”, indicator 2: “The range and appropriateness of teaching methods”. Under category “teacher knowledge”, indicator 4: “Ability of the teacher to contextualise knowledge with locally relevant examples”. Both indicators are rated on a scale from 1 to 4 but with no rubric nor examples.

Coded as 2
Philippines
Indicator 3: “Selects, develops, organizes, and uses appropriate teaching and learning resources, including ICT, to address learning goals”. Rated on a scale from 1 to 9 using a rubric. As there is a separate indicator for ICT use, the rubric of this indicator focuses on “extensive and multidisciplinary learning resources”, which implies the use of multiple representations.

CLASS
Under category “instructional support”, indicator 1: “Instructional learning formats”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “modalities”, specifically the “variety of modalities for presenting information”.

MQI
Under category “richness of the mathematics”, indicator 1: “Linking between representations”. Rated on a scale from “not present” to “high” using a rubric. This indicator requires “different representations of a mathematical idea or procedure”, implying variety.

UTOPI
Under category “mathematics/science content”, indicator 4.5: “Content abstraction”. Rated on a scale from 1 to 5 using a rubric. This indicator measures “multiple forms
of representation in science and mathematics classes include graphic, verbal, symbolic, visualisations…”.

Coded as 0
We cannot find evidence of instructional scaffolding in the following tools: Brunei, The Gambia, Guyana, Kenya (18), Lebanon, Liberia, Malawi, Pakistan (58), Tajikistan (29), Tajikistan (30), Vietnam, FFT, SCOPE, Stallings, TIPPS.

Sub-indicator 6.36 – Opportunities to practice applying concepts/procedures
Definition: The degree to which the teacher provides students with the opportunity to practice applying concepts, procedures, or skills, either independently or under supervision.

Coded as 1
Brunei
Under category “students’ achievements”, indicator 4: “application of knowledge, skills and understanding – + / –”.

The Gambia
Under category “student engagement and learning”, indicator 2: “Students were given appropriate tasks and teacher gave support to all students”. Rated on a scale from 1 to 5, however with no rubric nor examples.

Kenya (18)
In multiple instances of the tool, questions such as “what ______ strategies were modelled? – list of strategies”. For example, “what vocabulary strategies were modelled? – gesture, make a sentence, give definition, point to something, use a picture, give another word in English/MT or Kiswahili, not applicable, this section was skipped”. These are all different methods of practicing and applying vocabulary knowledge.

Kenya (20)
Under the category “lesson plan”, indicator 2.6: “Setting of homework (if appropriate) – unsatisfactory, satisfactory, good, very good”.

Malawi
Indicator 11: “teaching for effective learning”. To score a 3 or above on this indicator, teachers need to “set homework and other out-of-class activities to consolidate and extend students’ knowledge and understanding”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, hence we code it as one.

Pakistan (57)
Under category “evaluate during the lesson”, indicator 4: “Activity based teaching and learning”. To score above a 1 on this indicator, the lesson needs to include “an activity performed by students which is instructive and relevant to SLOs along with reading”. Although rated from 0 to 3, the rating is done progressively as a checklist. There is no rubric. Therefore, we code this as one.
Papua New Guinea (10)
Under category “do the following at the END of the observation”, observers examine students’ exercise books and look for a range of activities. These are indicators 9a through 9h, for example 9a “filling in missing words, sentence beginnings and endings, etc. – yes, no”, 9b “short sentences – yes, no”, and 9e “labelling things – yes, no”. This examination effectively looks for students’ opportunities to practice concepts.

Papua New Guinea (11)
Under category “do the following at the END of the observation”, observers examine students’ exercise books and look for a range of activities. These are indicators 9a through 9h, for example 9a “filling in missing words, sentence beginnings and endings, etc. – yes, no”, 9b “short sentences – yes, no”, and 9e “labelling things – yes, no”. This examination effectively looks for students’ opportunities to practice concepts.

Tajikistan (30)
Indicator 22: “Teacher gives students enough time to practice learning skill – yes, no, difficult to determine”.

Uganda (27)
Under category “quality of teaching and learning process”, indicator 4: “The extent to which learners apply theoretical knowledge during learning process”. Rated on a scale from 1 to 4 but with no rubric nor examples.

Uganda (28)
Under category “practical component of the lesson”, indicator 15: “The lesson included sufficient time for learners to practice e.g. make exercises, solve problems, conduct experiments…”. Rated on a scale from 1 to 4 but with no rubric nor examples.

SDI
Under category “teacher asking questions”, indicator 34: “The teacher asked learners to carry out a task which allowed them to demonstrate their understanding of what they had learned during the lesson – yes, no”.

Stallings
Indicator 4: “Practice & drill”. Coded as a yes or no question, with the observer indicating whether the whole class, a large group of students, a small group of students, or one student, are involved in the activity.

Coded as 2
Guyana
Under category “pedagogy”, indicators 5 and 11: “Learners have/use opportunities to demonstrate their new knowledge and/or skill related to concepts in the lesson”, and “opportunity is provided for further work on concepts outside of class”. Both indicators are rated on a scale from 0 to 4 using a rubric.

CLASS
Under category “instructional support”, indicator 2: “Content understanding”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “opportunity for practice of procedures and skills” at every rating level.

**SCOPE**
Under category “language and literacy instruction”, all indicators, “opportunities for [literacy learning type]”. Rated on a scale from 1 to 5 using a rubric. For example, indicator 9: “Opportunities for learning to decode and spell words”. The rubric measures the extent to which these opportunities are then “immediately applied in new contexts”.

**TIPPS**
Indicator 4B: “Teacher uses instructional materials and/or activities to facilitate exploration and further learning”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 4A, which is also rated on the same scale.

Coded as 0
We cannot find evidence of instructional scaffolding in the following tools:
Kenya (19), Lebanon, Liberia, Pakistan (58), Philippines, Tajikistan (29), Uganda (26), Vietnam, FFT, MQI, UTOP.

**Indicator 6.40 – Clear presentation and lesson structure**
Definition: The degree to which the teacher presents information in a clear and organized manner, stays on topic, keeps the lesson moving, or does not allow students to spend excessive time on one task.

Coded as 1
Brunei
This tool received a one under sub-indicator 6.41.

The Gambia
Under category “presentation/lesson delivery”, indicators 1 and 5: “Voice was clear and audible”, and “chalkboard work was clear and legible”. Both indicators are rated on a scale from 1 to 5 but with no rubric or examples. Additionally, under category “classroom management”, indicator 2: “Gave clear instructions”. Also rated on a scale from 1 to 5 but with no rubric nor examples.

Guyana
Under category “classroom and student management”, indicator 2: “Instructions are clearly communicated”. Rated on a scale from 0 to 4, but with no rubric nor examples.

Kenya (18)
“Based on the lesson, how prepared was the teacher? – unprepared, prepared, well prepared”.
Kenya (19)
“Based on the lesson, how prepared was the teacher? – unprepared, prepared, well prepared”.

Kenya (20)
“Was the teacher prepared? – yes, somehow, no”. Additionally, “teacher explains material accurately and clearly – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon

Malawi
Indicator 11: “Teaching for effective learning”. To score a 3 or above on this indicator, teachers need to “present lessons clearly, and ensure their instructions and explanations are easy to understand”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, therefore we code it as one.

Pakistan (57)
Under category “evaluate after the lesson”, indicator 8: “Interaction with students”. To score a 1 or above on this indicator teachers need to be “clear and audible”. Although rated from 0 to 3, the rating is done progressively as a checklist. There is no rubric. Therefore, we code this as one.

Pakistan (58)
Under category “the teacher”, indicator 3: “Was the teacher’s voice clear and easily heard? – yes, no, to some extent”. Additionally, under category “classroom management”, indicator 3: “Were the instructions for the activities clear? – yes, no, to some extent”.

Tajikistan (30)
Indicator 18: “Teacher provides clear instructions to students on how to use agreed algorithm to solve learning problem – a) yes, b) no, c) partly”.

Uganda (26)
Indicator 2: “Ability to communicate clearly”. Rated on a scale of three from “needs development” to “well developed”, but with no rubric nor examples.

Uganda (27)
Under category “Quality of teaching and learning process”, indicator 3: “clarity of teacher presentation”. Rated on a scale from 1 to 4, but with no rubric nor examples.

Uganda (28)
Under category “lesson structure and content”, indicator 5: “The objectives of the lesson and the relation with previous,next lessons were clearly explained”. Rated on a scale from 1 to 4, but with no rubric nor examples.
This tool received a one under sub-indicator 6.42.

Coded as 2

Philippines

Indicator 9: “Uses effective verbal and non-verbal classroom communication strategies to support learner understanding, participation, engagement, and achievement”. Rated on a scale from 1 to 9 using a rubric. This indicator includes a measure of the degree to which the “teacher uses clear communication”.

Tajikistan (29)

Under category “lesson strategies and delivery”, indicator II: “Clarity of learning objectives”. Rated on a scale from 1 to 7 using a rubric. This indicator includes measures of the degree to which the teacher provides “clear instructions to students”.

Vietnam

Indicator 13: “Routines”. This indicator includes a measure of “clear instructions” given by the teacher. Rated on a scale of 3 from “basic” to “advanced”, using a rubric.

CLASS

This tool received a two under sub-indicator 6.42.

FFT

Under category “instruction”, indicator 3a: “Communicating with students”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “explanations of content”, and the degree to which “the explanations are clear”.

SCOPE

Under category “”, indicator 6: “Manages reading and writing instruction”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of whether “clear instructions about what learners are expected to do are given”.

UTOP

This tool received a two under both sub-indicator 6.41 and sub-indicator 6.42.

Coded as 0

We cannot find evidence of clear presentation and lesson structure in the following tools: Liberia, Papua New Guinea (10), Papua New Guinea (11), Stallings, TIPPS.

Sub-indicator 6.41 – Clear learning objectives
Definition: The degree to which the teacher clearly communicates learning objectives or re-orient students to the objectives; students are aware of the lesson’s purpose.

Coded as 1

Brunei
Under category “teaching”, indicator 4.1: “Establishing and sharing clear learning objectives - + / – ”.

The Gambia
Under category “preparation and planning”, indicator 3: “Objectives are clearly stated in the lesson plan”. Rated on a scale from 1 to 5 but with no rubric or examples.

Kenya (20)
Under category “scheme of work”, indicator 1.1: “Clarity of learning objectives – unsatisfactory, satisfactory, good, very good”. Additionally, “teacher clearly states objectives and activities to be covered and refers to them throughout the lesson – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon
Under category “lesson plan”, indicator 1: “Clear competencies and learning objectives – remarks”.

Malawi
Indicator 11: “Teaching for effective learning”. To score a 4 on this indicator, teachers need to “explain to students what they are going to learn”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, therefore we code it as one.

Pakistan (57)
Under category “evaluate during the lesson”, indicator 3: “SLO (Taleemi calendar)”. To score a 2 or above on this indicator teachers need to state “clear and specific SLOs”. Although rated from 0 to 3, the rating is done progressively as a checklist. There is no rubric. Therefore, we code this as one.

Pakistan (58)
Under category “the objective(s)”, indicator 3: “Were the objectives clear? – yes, no, to some extent”.

Tajikistan (30)
Indicator 15: “Teacher presented and discussed the lesson’s learning objectives – a) not presented, b) presented only, c) presented and discussed, d) discussed only, e) difficult to determine”.

Uganda (27)
Under category “quality of teacher’s planning”, indicator 3: “Level of clarity in the statement of learning outcomes/competencies”. Rated on a scale from 1 to 4 but with no rubric nor examples. Additionally, under category “quality of teaching and learning process”, indicator 1: “Clarity of purpose of lesson”. Also rated on a scale from 1 to 4 but with no rubric nor examples.

Uganda (28)
Under category “lesson structure and content”, indicator 4: “The objectives of the lesson and the relation with previous/next lessons were clearly explained”.Rated on a scale from 1 to 4, but with no rubric nor examples.

**Coded as 2**

**Guyana**

Under category “planning”, indicator 1: “Objectives are clearly written to focus on learner outcomes”. Rated on a scale from 0 to 4 using a rubric.

**Tajikistan (29)**

Under category “lesson strategies and delivery”, indicator II: “Clarity of learning objectives”. Rated on a scale from 1 to 7 using a rubric.

**CLASS (K-3)**

Under category “instructional learning formats”, indicator 3: “Clarity of learning objectives”. Rated on a scale from 1 to 7 using a rubric.

**FFT**

Under category “planning and preparation”, indicator 1c: “Setting instructional outcomes”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “clarity” of learning outcomes. Additionally, under category “instruction”, indicator 3a: “Communicating with students”. Also rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “expectations for learning”, and the degree to which “the goals for learning are communicated clearly to students”.

**MQI**

Under category “errors and imprecision”, indicator 3: “Lack of clarity in presentation of mathematical content”.

**SCOPE**

Under category “”, indicator 6: “Manages reading and writing instruction”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of whether the teacher has clear “lesson objectives”, and whether “instruction is aligned with the objectives of the lesson”.

**UTOPI**

Under category “lesson structure”, indicator 2.1: “Lesson sequence”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which the lesson had “a clear sense of purpose and clearly stated objectives”.

**Coded as 0**

We cannot find evidence of clear learning objectives in the following tools: Kenya (18), Kenya (19), Liberia, Papua New Guinea (10), Papua New Guinea (11), Philippines, Uganda (26), Vietnam, CLASS, MQI, SDI, Stallings, TIPPS.
Sub-indicator 6.42 – Clear lesson sequence

Definition: The degree to which each topic or learning exercise has a clear introduction and conclusion, and/or is sequenced logically.

Coded as 1

The Gambia
Under category “presentation/lesson delivery”, indicator 6: “Lesson was structure well; with a clear introduction and conclusion”. Rated on a scale from 1 to 5 but with no rubric or examples.

Kenya (20)
Under category “scheme of work”, indicator 1.2: “Sequencing of lessons – unsatisfactory, satisfactory, good, very good”.

Lebanon

Pakistan (58)
Under category “classroom management”, indicator 6: “Was there a clear introduction and conclusion to the lesson? – yes, no, to some extent”.

Uganda (27)
Under category “quality of teacher’s planning”, indicator 3: “The logical description of the structure of the lesson and teaching methods to be used”. Rated on a scale from 1 to 4 but with no rubric nor examples.

Uganda (28)
Under category “lesson structure and content”, indicator 7: “key points were summarised during and/or at the end of the lecture”.

SDI
Under category “introducing and summarizing the lesson”, indicators 40 and 41: “Did the teacher introduce the lesson at the start of the class? – yes, no, don’t know”, and “did the teacher summarize the lesson at the end of the class? – yes, no, don’t know”.

Coded as 2

Guyana
Under category “planning”, indicator 3: “Content is organised and sequenced logically”. Rated on a scale from 0 to 4 using a rubric.

FFT
Under category “planning and preparation”, indicator 1e: “Designing coherent instruction”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “lesson and unit structure”, the degree to which “teachers produce clear and sequenced lesson and unit structures”.
SCOPE
Under category “”, indicator 6: “Manages reading and writing instruction”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of whether “the teacher follows a regular pattern of instructional activities that loosely build on skills and concepts previously taught in an appropriately sequenced manner”.

UTOP
Under category “lesson structure”, indicator 2.1: “Lesson sequence”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the “sequence of the lesson as a whole, and whether it had appropriate and reasonable engagement, learning, and wrap up activities”.

Coded as 0
We cannot find evidence of clear introduction and conclusion the following tools:
Brunei, Kenya (18), Kenya (19), Liberia, Malawi, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (29), Uganda (26), Vietnam, CLASS, MQI, Stallings, TIPPS.

Indicator 6.50 – Language development and modelling
Definition: The degree to which teacher uses language accurately, encourages students to use language correctly, and assists them in the process.

Coded as 1
The Gambia
Under category “presentation/lesson delivery”, indicator 2: “Demonstrated a good command of the language of instruction”.

Kenya (18)
“Did the teacher pronounce the sound correctly? – yes, no”, “was the teacher able to take apart, put together or change sounds? – yes, no”.

Lebanon

Pakistan (58)
Under category “personal qualities”, indicator f: “Use of English”. Rated on a scale of 3 from “acceptable or competent” to “unacceptable” but with no rubric nor examples.

Papua New Guinea (10)

Papua New Guinea (10)
Under category “teacher’s competency in the language of instruction”, indicators 4a through 4c: “Teacher makes errors in spoken English? – yes, no”, “errors in teachers’ written text in English on the black board? – yes, no”, and “errors in teachers’ written text in English on wall charts? – yes, no”

SDI

This tool received a one under sub-indicator 6.51.

Coded as 2

Philippines

Indicator 8: “Displays proficient use of Mother Tongue, Filipino and English to facilitate teaching and learning”. Rated on a scale from 1 to 8 using a rubric.

CLASS (K-3)

Under category “instructional support”, indicator 3: “Language modelling”. Rated on a scale from 1 to 7 using a rubric.

FFT

Under category “instruction”, indicator 3a: “Communicating with students”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “use of oral and written language”, measured because “for many students, their teachers’ use of language represents their best model of both accurate syntax and a rich vocabulary”.

MQI

Under category “errors and imprecision”, indicator 2: “Imprecision in language or notation”. Rated on a scale of 4, from “not present” to “high”, using a rubric.

SCOPE

Under category “language and literacy instruction”, indicator 7: “Opportunities for oral language development”, Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which the “teacher uses language that learners understand or explains unfamiliar language” at every rating level.

TIPPS

Indicator 8B: “The teacher models high quality language expression to advance pupil understanding and use of language”. Rated on a scale from “accurate” to “very accurate”, against its converse, indicator 8A, which is also rated on the same scale.

Coded as 0

We cannot find evidence of clear introduction and conclusion the following tools:

Brunei, Guyana, Kenya (19), Kenya (20), Liberia, Malawi, Pakistan (57), Stallings, Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, CLASS, UTOP.

Sub-indicator 6.51 – Use of local language

Definition: The degree to which teacher uses local language to help students grasp complex concepts in English (or the primary language of instruction).
Papua New Guinea (10)
Under category “languages used in the classroom”, indicators 3a and 3c: “Language(s) of instruction used by teachers – Kuanua, English & Kuanua, Tok Pisin & Kuanua, Other (specify)”, and “language(s) written on the black board to teach content – “Kuanua, English & Kuanua, Tok Pisin & Kuanua, Other (specify)”.

Papua New Guinea (11)
Under category “languages used in the classroom”, indicators 3a and 3c: “Language(s) of instruction used by teachers – English, Tok Pisin, English & Tok Pisin, English and vernacular, Tok Pisin & vernacular, vernacular (specify)”, “language(s) written on the black board to teach content – English, Tok Pisin, English & Tok Pisin, English and vernacular, Tok Pisin & vernacular, vernacular (specify)”.

SDI
Under category “language”, indicator 44: “Did the teacher use the local language of instruction? (language other than English) – yes, no”

Coded as 2
Philippines
Indicator 8: “Displays proficient use of Mother Tongue, Filipino and English to facilitate teaching and learning”. Rated on a scale from 1 to 8 using a rubric.

Coded as 0
We cannot find evidence of clear introduction and conclusion the following tools: Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, CLASS, FFT, MQI, SCOPE, Stallings, TIPPS, UTOP.

Indicator 6.60 – Lesson facilitation and discourse
This is not a stand-alone indicator, meaning we only code this as one if one of the sub-indicators is also coded as a one.

Coded as 1
Brunei
This tool received a one for sub-indicator 6.64

The Gambia
This tool received a one for sub-indicator 6.61, sub-indicator 6.63, and sub-indicator 6.65.

Kenya (18)
This tool received a one for sub-indicator 6.61, sub-indicator 6.63, and sub-indicator 6.65.
Kenya (20)
This tool received a one for both sub-indicator 6.61 and sub-indicator 6.65.

Lebanon
This tool received ones for all sub-indicators.

Malawi
This tool received a one for sub-indicator 6.64.

Pakistan (57)
This tool received a one for sub-indicator 6.62.

Pakistan (58)
This tool received a one for both sub-indicator 6.63 and sub-indicator 6.64.

Papua New Guinea (10)
This tool received a one for both sub-indicator 6.61 and sub-indicator 6.65.

Papua New Guinea (11)
This tool received a one for both sub-indicator 6.61 and sub-indicator 6.65.

Tajikistan (30)
This tool received a one for both sub-indicator 6.62 and sub-indicator 6.65.

Uganda (26)
This tool received a one for sub-indicator 6.63.

Uganda (27)
This tool received a one for both sub-indicator 6.61 and sub-indicator 6.64.

Uganda (28)
This tool received a one for both sub-indicator 6.61 and sub-indicator 6.64.

SDI
This tool received a one for sub-indicator 6.61.

Stallings
This tool received a one for both sub-indicator 6.61 and sub-indicator 6.62.

**Coded as 2**
Guyana
This tool received a two for sub-indicator 6.61, sub-indicator 6.63, and sub-indicator 6.64.

Philippines
This tool received twos for all sub-indicators.
Tajikistan (29)
This tool received a two for sub-indicator 6.61, sub-indicator 6.62, sub-indicator 6.64, and sub-indicator 6.65.

Vietnam
This tool received a two for sub-indicator 6.62, sub-indicator 6.64, and sub-indicator 6.65.

CLASS
This tool received twos for all sub-indicators.

FFT
This tool received twos for all sub-indicators.

MQI
This tool received a two for sub-indicator 6.65.

SCOPE
This tool received a two for sub-indicator 6.61, sub-indicator 6.62, sub-indicator 6.63, and sub-indicator 6.65.

TIPPS
This tool received a two for sub-indicator 6.61, sub-indicator 6.62, sub-indicator 6.64, and sub-indicator 6.65.

UTOP
This tool received a two for sub-indicator 6.62, sub-indicator 6.63, sub-indicator 6.64, and sub-indicator 6.65.

Coded as 0
We cannot find evidence of lesson facilitation and discourse in the following tools:
Kenya (19), Liberia.

Sub-indicator 6.61 — A variety of teacher-directed instructional strategies
Definition: The degree to which the teacher uses a variety of teacher directed strategies, approaches, or materials (for example reading aloud, watching a video, singing a song, etc.) to actively engage all students in the lesson.

Coded as 1
The Gambia
This is a time snapshot tool that records classroom activity every five minutes, on list of potential activities. Items on the list include “teacher presenting or explaining” and “question and answer, discussion”. These are varied teacher-directed instructional strategies.

Kenya (18)
In multiple instances of the tool, questions such as “what ______ strategies were modelled? – list of strategies”. For example, “what vocabulary strategies were modelled? – gesture, make a sentence, give definition, point to something, use a picture, give another word in English/MT or Kiswahili, not applicable, this section was skipped”. These are all different teacher-directed instructional strategies.

Kenya (20)
Under “scheme of work”, indicator 1.3: “Range of teaching and learning activities – unsatisfactory, satisfactory, good, very good”.

Lebanon
Under category “instructions”, all indicators. This category lists a number of instructional strategies, some of them teacher directed. For example, “class discussion”, “lecturing”, “explanation and note taking”.

Papua New Guinea (10)
Under category “teaching”, indicator 6i: “Varied reading activities during the lesson (e.g. spelling, verbal reporting, general discussions, reading with teachers…) – yes, no”.

Papua New Guinea (11)
Under category “teaching”, indicator 6i: “Varied reading activities during the lesson (e.g. spelling, verbal reporting, general discussions, reading with teachers…) – yes, no”.

Uganda (27)
Under category “quality of teaching and learning process”, indicator 2: “The range and appropriateness of teaching methods”. Rated on a scale from 1 to 4 but with no rubric nor examples.

Uganda (28)
Under category “ATL methods, techniques and teaching aids”, all indicators. These demonstrate a list of instructional strategies, some of them teacher directed. For example, “brainstorming”, “demonstration”, “research”, and “storytelling and shared writing”. If the techniques were used, they were rated on a scale from 1 to 4 but with no rubric nor examples.

Coded as 2
Guyana
Under category “planning”, indicator 6: “Teaching strategies are appropriate to learners and objectives”. Additionally, under category “pedagogy”, indicator 1: “Teaching/learning strategies enhance the key concepts”. Both indicators are rated on a scale from 0 to 4 using a rubric.

Philippines
Indicator 2: “Uses a range of teaching strategies to develop critical and creative thinking, as well as other high-order thinking skills”. Rated on a scale from 1 to 9 using a rubric.

Tajikistan (29)
Under category “lesson strategy and delivery”, indicator III: “Variety of resources and materials”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the “variety of strategies including auditory, visual, and movement” the teacher uses.

CLASS
Under category “instructional support”, indicator 1: “Instructional learning formats”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “variety of modalities, strategies and materials”.

FFT
Under category “planning and preparation”, indicator 1e: “Designing coherent instruction”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “learning activities”, and pointers of this indicator include “opportunities for student choice” and “use of varied resources”. Additionally, under category “instruction”, indicator 3e: “Demonstrating flexibility and responsiveness. Also rated on a scale from 1 to 4 using a rubric. This indicator includes “persistence”, the degree to which “teachers seek alternate approaches to help their students be successful”.

SCOPE
Under category “language and literacy instruction”, all indicators, “opportunities for [literacy learning type]”. Rated on a scale from 1 to 5 using a rubric. These measure the extent to which different activities are used in the classroom, including “writing instruction”, “reading fluency”, and “oral language development”. Therefore, these indicators also measure the variety of instructional strategies in the classroom.

SDI
This is a time snapshot tool that records the “teacher’s activity” every five minutes, on a checklist. Items on the checklist include “teacher interacts with all children as a group”, “teacher reads or lectures to the pupils”, “teacher leads kinaesthetic group learning activity”. These are varied teacher-directed instructional strategies.

Stallings
This is a time snapshot tool that records classroom activity every few minutes, on list of potential activities. Items on the list include “practice & drill”, “demonstration/lecture”, “discussion/debate/questions and answers”. These are varied teacher-directed instructional strategies.

TIPPS
Indicator 3B: “Teacher uses instructional strategies to aid students in complex thinking or problem solving”. Rated on a scale from “accurate” to “very accurate”, against its converse, indicator 3A, which is also rated on the same scale. Examples of possible instructional strategies include “comparisons and contrasts, story-telling, and/or problem-solving activities”.

**Coded as 0**
We cannot find evidence of teacher-directed instructional strategies in the following tools: Brunei, Kenya (19), Liberia, Malawi, Tajikistan (30), Uganda (26), MQI, UTOP.
Sub-indicator 6.62 – Discussion facilitation
Definition: The degree to which the teacher and students facilitate back and forth discussion.

Coded as 1
Lebanon

Pakistan (57)
Under category “evaluate after the lesson”, indicator 8: “interaction with students”. To score a 3 on this indicator teachers need to “interact with all students in class, creating a dialogue with all students”. Although rated from 0 to 3, the rating is done progressively as a checklist. There is no rubric. Therefore, we code this as one.

Tajikistan (30)
Indicator 17: “What below activities were applied by teachers to guide students to solve the problem (mark all that apply) – a) encourage students to find decisions independently, b) encourage students to discuss different ways to solve issue, c) students define the agreed way to solve the [issue], d) teacher demonstrates/models skills or sample how to solve the learning problem, e) students compare and discuss their own way with that demonstrated by teacher and finalize the way to solve the learning problem (algorithm), f) none of the listed”. Responses b, c, and e show discussion facilitation.

Stallings
Indicator 3: “Discussion/debate/questions and answers”. Coded as a yes, no question, with the observer indicating whether the whole class, a large group of students, a small group of students, or one student, are involved in the activity.

Coded as 2
Philippines
Indicator 9: “Uses effective verbal and non-verbal classroom communication strategies to support learner understanding, participation, engagement, and achievement”. Rated on a scale from 1 to 9 using a rubric. To score highly on this indicator, teachers need to “successfully facilitate student discussion”.

Tajikistan (29)
Under category “lesson strategy and delivery”, indicator I: “Effective facilitation”. Rated from 1 to 7 using a rubric. To score highly on this indicator, teacher “engages most students in the discussions”, and “embeds questions in the discussion”. Therefore, this indicator includes elements of discussion facilitation.

Vietnam
Indicator 20: “Problem solving skills”. Rated on a scale of 3 from “basic” to “advanced”. This indicator measures the extent to which the teacher “uses discussion and activities that encourage analysis”.

CLASS
Under category “instructional support”, indicator 5: “Instructional dialogue”. Rated on a scale from 1 to 7 using a rubric. This indicator includes measures of “facilitation strategies” and “distributed talk”.

FFT
Under category “instruction”, indicator 3b: “Using questioning and discussion techniques”. Rated on a scale from 1 to 4 using a rubric.

SCOPE
Under category “language and literacy instruction”, indicator 7: “Opportunities for oral language development”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the use of “discussion” to develop language.

TIPPS
Indicator 3B: “Teacher uses instructional strategies to aid students in complex thinking or problem solving”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 3A, which is also rated on the same scale. This indicator and its converse measure whether the teacher “initiates discussion”.

UTOP
Under category “mathematics/science content”, indicator 4.5: “Content Abstraction: Elements of mathematical/scientific abstraction were used appropriately”. This indicator includes measures of “explicit discussion” at each rating level.

Coded as 0
We cannot find evidence of discussion facilitation in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Liberia, Malawi, Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Uganda (26), Uganda (27), Uganda (28), MQI, SDI.

Sub-indicator 6.63 – Grade appropriate lesson materials
Definition: Lesson materials and teacher’s language are appropriate for the grade level, and are consistent with students’ needs and abilities (from CLASS).

Coded as 1
The Gambia
Under category “presentation/lesson delivery”, indicator 10: “The lesson was appropriate for the class level”. Rated on a scale from 1 to 5 but with no rubric nor examples.

Kenya (18)
“Was the teacher able to teach grammar to the level of pupils understanding? – yes, no”

Lebanon
Under category “instructions”, indicator 5: “Suitable according to the students’ levels and need – remarks”.

Pakistan (58)
Under the category “the pupils”, indicator 1: “were the activities appropriate for the students? – yes, no, to some extent”.

Uganda (26)
Under category “facilitation”, indicator 6: “Use of language which is at the level of participants”.

Coded as 2
Guyana
Under category “planning”, indicators 2 and 4: “Objectives are in keeping with the context of the curriculum and the developmental stage of the learners”, and “concepts are developmentally appropriate in scope and depth”. Rated on a scale from 0 to 4 using a rubric.

Philippines
Indicator 13: “Uses differentiated, developmentally appropriate learning experiences to address learners’ gender, needs, strengths, interests and experiences”. Rated on a scale from 1 to 9 using a rubric. We code this here due to the emphasis on “developmentally appropriate learning”.

CLASS (K-3 only)
Under category “emotional support”, indicator 2: “Teacher sensitivity”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “appropriate activities”, which shows the extent to which the “teacher provides activities or speaks at levels consistent with student needs and abilities”.

FFT
Under category “planning and preparation”, indicator 1b: “Demonstrating knowledge of students”. Rated on a scale from 1 to 4 using a rubric. This indicator includes measures of “knowledge of child and adolescent development”, and “knowledge of the learning process”.

SCOPE
Under category “classroom culture”, indicator 5: “Classroom materials”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the degree to which materials are “appropriate to the curriculum” and “appropriate for the grade level”.

UTOP
Under category “Mathematics/science content”, indicator 4.1: “Content significance: the mathematics or science content chosen was significant, worthwhile, and
developmentally appropriate for this course”. Rated on a scale from 1 to 5 using a rubric.

Coded as 0
We cannot find evidence of grade appropriate lesson materials in the following tools: Brunei, Kenya (19), Kenya (20), Liberia, Malawi, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (29), Tajikistan (30), Uganda (27), Uganda (28), Vietnam, CLASS, MQI, SDI, Stallings.

Sub-indicator 6.64 – Teacher flexibility and tailoring to differential student needs
Definition: The degree to which the teacher tailors or individualizes support to students or demonstrates student needs flexibility with the lesson (such as structure, focus, topic, or timing) to accommodate students’ ability levels or understanding.

Coded as 1
Brunei
Under category “teaching”, indicator 3: “Suits work to students’ individual needs – + / –”.

Lebanon
Under category “lesson delivery”, indicator 10: “Varying activities according to the students’ abilities, learning styles and needs – remarks”.

Malawi
Indicator 11: “Teaching for effective learning”. To score a 4 on this indicator, teachers need to “adjust their teaching method to class size, facilities available, what students need to learn and how well they understand”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, therefore we code it as one.

Pakistan (58)
Under category “personal qualities”, indicator h: “Flexibility”. Rated on a scale of 3, from “acceptable or competent” to “unacceptable”, but with no rubric nor examples.

Uganda (27)
Under category “quality of teacher’s planning”, indicator 6: “The extent to which lessons reflect the diverse special needs of the learners”. Rated on a scale from 1 to 4 but with no rubric nor examples.

Uganda (28)
Under category “teacher-student interactions and participation”, indicator 13: “Teacher provided special assistance to learners with difficulties”. Rated on a scale from 1 to 4 but with no rubric nor examples.

Coded as 2
Guyana
Under category “pedagogy”, indicator 9: “Teaching/learning activities cater to learners’ differences”. Rated on a scale from 0 to 4 using a rubric.
Philippines
Indicator 13: “Uses differentiated, developmentally appropriate learning experiences to address learners’ gender, needs, strengths, interests and experiences”. Rated on a scale from 1 to 9 using a rubric.

Tajikistan (29)
Under category “classroom climate”, indicator III: “Teacher sensitivity”. This indicator includes a measure of the degree to which the teacher is “responsive to students and matches her support to their needs and abilities”. Rated on a scale from 1 to 7 using a rubric.

Vietnam
Indicator 12: “Flexibility and student focus”. Rated on a scale of three from “basic” to “advanced” using a rubric.

CLASS
Under category “emotional support”, indicator 3: “Regard for student perspectives”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of teachers’ “flexibility and student focus”.

FFT
Under category “instruction”, indicator 3e: “Demonstrating flexibility and responsiveness”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “lesson adjustment”.

TIPPS
Indicator 4B: “Teacher sensitivity and responsiveness to students’ needs and learning”. Rated on a scale from “accurate” to “very accurate”, against its converse, indicator 4A, which is also rated on the same scale. This includes the teacher “consistently assessing student needs”.

UTOP
Under category “implementation”, indicator 3.3: “Implementation modification”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which the teacher “modified the lesson appropriately when formative assessment demonstrated that students did not understand”.

Coded as 0
We cannot find evidence of teacher flexibility in the following tools:
Kenya (18), Kenya (19), Kenya (20), Liberia, Papua New Guinea (10), Papua New Guinea (11), Tajikistan (30), Uganda (26), MQI, SDI, Stallings.

Sub-indicator 6.65 – Encouraging peer-to-peer interaction or group work
Definition: The degree to which teacher engages students to work together on the task at hand.

Coded as 1
Brunei
Under category “students’ learning”, Indicator 8: “Cooperative learning – + / –”.

The Gambia
Under category “time on task – please tick the main type of activity during each 5-minute period”, “student activity, individual or group”.

Kenya (18)
“Who read the story? (mark all that apply) – teacher, whole class (echo/choral), partners/groups, individual, this section was skipped”.

Kenya (20)
“Teacher uses paired or group work to encourage active participation of students – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon

Papua New Guinea (10)
Under category “teaching”, indicators 6f and 6g: “Students assisting peers to read (buddy reading) – yes, no”, and “students reading to each other in groups – yes, no”.

Papua New Guinea (11)
Under category “teaching”, indicators 6f and 6g: “Students assisting peers to read (buddy reading) – yes, no”, and “students reading to each other in groups – yes, no”.

Tajikistan (30)
Indicator 17: “What below activities were applied by teachers to guide students to solve the problem (mark all that apply) – a) encourage students to find decisions independently, b) encourage students to discuss different ways to solve issue, c) students define the agreed way to solve the [issue], d) teacher demonstrates/models skills or sample how to solve the learning problem, e) students compare and discuss their own way with that demonstrated by teacher and finalize the way to solve the learning problem (algorithm), f) none of the listed”. Responses b and e show peer-to-peer interactions. Additionally, indicator 24: “During the lesson students were involved in (mark all that apply) – a) individual work, b) pair work, c) small group work, d) students answer teachers questions, e) students give questions and answer on the other students questions, f) other”. Answers b, c, and e, show peer-to-peer interaction.

Uganda (28)
Under category “ATL methods, techniques and teaching aids”, indicators 3 and 9: “group work – technique/method used?” and “pair work and peer-to-peer explanations – technique/method used?” If the techniques were used, they were rated on a scale from 1 to 4 but with no rubric nor examples.
Coded as 2
Philippines
Indicator 10: “Maintains supportive learning environments that nurture and inspire learners to participate, cooperate, and collaborate in continued learning”. Rated on a scale from 1 to 9 using a rubric. This indicator measures the extent of “collaborative learning” and “group learning activities”.

Tajikistan (29)
Under category “lesson strategies and delivery”, indicator IV: “Effective use of group work”. Rated on a scale from 1 to 7 using a rubric.

Vietnam
Indicator 5: “Examination of activity segments”. Rated on a scale of 3 from “basic” to “advanced” using a rubric. Activity segment types include “group activity” and “pair activity”. Additionally, indicator 15: “Teamwork/cooperative learning”. Also rated on a scale of 3, from “basic” to “advanced” using a rubric.

CLASS
Under category “emotional support”, indicator 3: “Regard for student perspectives”. This includes a measure of “meaningful peer interactions”. Rated from 1 to 7 using a rubric.

FFT
Under category “planning and preparation”, indicator 1e: “Designing coherent instruction”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “instructional grouping”. Additionally, under category “instruction”, indicator 3c: “Engaging students in learning”. Also rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of the “grouping of students”. Both of these indicators assume the presence of peer-to-peer and group work.

MQI
Under category “common core aligned student practices”, indicator 3: “Students communicate about the mathematics of the segment”. Rated on a scale of 4 from “not present” to “high”. This indicator measures peer-to-peer interactions in “whole-group or small group” settings, at each rating level.

SCOPE
Under category “classroom culture”, indicator 2: “Effective grouping strategies”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the degree to which teachers use “whole group, small group, and cooperative learning groups”.

TIPPS
Indicator 1B: “Teacher creates opportunities for cooperative learning, so pupils are able to learn from one-another”. Rated on a scale from “accurate” to “very accurate”, against its converse, indicator 1A, which is also rated on the same scale.

UTOP
Under category “classroom environment”, indicator 1.2: “Classroom interactions: interaction reflected collegial working relationships among students”. Rated on a scale from 1 to 5 using a rubric. “This indicator captures how well the teacher has worked with the students on developing group-work ethics and skills that create and promote an environment of active collaboration.”

Coded as 0
We cannot find evidence of peer-to-peer interaction in the following tools:
Guyana, Kenya (19), Liberia, Malawi, Uganda (26), Pakistan (57), Pakistan (58), Uganda (27), SDI, Stallings.

Indicator 7.10 – Critical thinking
This is not a stand-alone indicator, meaning we only code this as one if one of the sub-indicators is also coded as a one.

Coded as 1
Brunei
This tool received a one under sub-indicator 7.13.
The Gambia
This tool received a one under sub-indicator 7.11.
Kenya (20)
This tool received a one under sub-indicator 7.11.
Lebanon
This tool received a one under sub-indicator 7.11.
Malawi
This tool received a one under sub-indicator 7.13.
Tajikistan
This tool received a one under sub-indicator 7.12.
SDI
This tool received a one under sub-indicator 7.11.

Coded as 2
Philippines
This tool received a two under both sub-indicator 7.11 and sub-indicator 7.12.
Tajikistan (29)
This tool received a two under both sub-indicator 7.11 and sub-indicator 7.13.
Vietnam
This tool received a two under both sub-indicator 7.11 and sub-indicator 7.12.
CLASS
This tool received a two under both sub-indicator 7.11 and sub-indicator 7.12.

FFT
This tool received a two under all sub-indicators.

MQI
This tool received a two under sub-indicator 7.13.

SCOPE
This tool received a two under both sub-indicator 7.11 and sub-indicator 7.12.

TIPPS
This tool received a two under sub-indicator 7.11.

UTOP
This tool received a two under both sub-indicator 7.11 and sub-indicator 7.12.

Coded as 0
We cannot find evidence of critical thinking in the following tools:
Guyana, Kenya (18), Kenya (19), Liberia, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Uganda (26), Uganda (27), Uganda (28), Stallings.

Sub-indicator 7.11 – Thinking questions
Definition: The degree to which the teacher asks thinking questions, which are open-ended questions or those that require reasoning, explanation, or generalization and have more than one correct answer (TEACH).

Coded as 1
The Gambia
Under category “student engagement and learning”, indicator 3: “teacher asked questions to promote students’ critical thinking and check understanding”. Rated on a scale from 1 to 5 but with no rubric or examples.

Kenya (20)
“Teacher asks open-ended questions – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon
Under category “lesson deliver”, indicator 11: “Varying the types of questions (literal, inferential, analytical, etc.) – remarks”.

SDI
Under category “teacher asking questions”, indicator 36: “Did the teacher ask questions which required learners to use their creativity and imagination? – yes, no”. 
Coded as 2
Philippines
Indicator 2: “Applies a range of teaching strategies to develop critical and creative thinking, as well as other high-order thinking skills”. Rated on a scale from 1 to 9 using a rubric. This indicator refers to both “questions and activities”.

Tajikistan (29)
Under category “lesson strategies and delivery”, indicator I: “Effective facilitation”. Rated on a scale from 1 to 7 using a rubric. This indicator measures the degree to which “the teacher utilizes questioning techniques to enhance students learning”. It is about “open-ended questions to encourage students to think deeper”.

Vietnam
Indicator 8: “Use of enquiry mode”. Rated on a scale of 3 from “basic” to “advanced”, using a rubric. This indicator measures the degree to which teachers and students use “questions and counter-questions to develop understanding”. Additionally, indicator 20: “Problem solving skills”. Also rated on a scale of 3 from “basic” to “advanced”, using a rubric. This indicator measures the degree to which the “teacher often uses discussions and activities that encourage analysis and reasoning”.

CLASS
Under category “instructional support”, indicator 3: “Analysis and enquiry”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “facilitation of higher-order thinking”. This indicator refers to both “discussions and activities”.

FFT
Under category “planning and preparation”, indicator 1e: “Designing coherent instruction”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “activities that represent high-level thinking”.

SCOPE
Under category “language and literacy instruction”, indicator 7: “Opportunities for oral language development”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which “the teacher engages learners in extended conversations that support intellectual development (e.g., predictions, analysis, problem-solving, and reflection) and content knowledge”.

TIPPS
Indicator 9B: “Teacher asks open-ended questions and close-ended questions to facilitate deeper learning”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 9A, which is also rated on the same scale. Additionally, indicator 3B: “Teacher uses instructional strategies to aid students in complex thinking or problem solving”.

UTOPI
Under category “implementation”, indicator 3.1: “Implementation Questioning”. This indicator includes a measure of the degree to which “the teacher used questioning
strategies to encourage participation, check on skill development, and facilitate intellectual engagement and productive interaction with students about important science and mathematics content and concepts”.

**Coded as 0**

We cannot find evidence of thinking questions in the following tools:
Brunei, Guyana, Kenya (18), Kenya (19), Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), MQI, Stallings.

*Sub-indicator 7.12 – Thinking tasks*
Definition: The teacher provides thinking tasks that require summarizing, comparing, contrasting, and/or analyzing (TEACH).

**Coded as 1**

Tajikistan (30)
Indicator 17: “What below activities were applied by teachers to guide students to solve the problem (mark all that apply) – a) encourage students to find decisions independently, b) encourage students to discuss different ways to solve issue, c) students define the agreed way to solve the [issue], d) teacher demonstrates/models skills or sample how to solve the learning problem, e) students compare and discuss their own way with that demonstrated by teacher and finalize the way to solve the learning problem (algorithm), f) none of the listed”. Responses a, b, c, and e are thinking tasks.

**Coded as 2**

Philippines
Indicator 2: “Applies a range of teaching strategies to develop critical and creative thinking, as well as other high-order thinking skills”. Rated on a scale from 1 to 9 using a rubric. This indicator refers to both “questions and activities”.

Vietnam
Indicator 20: “Problem solving skills”. Rated on a scale of 3 from “basic” to “advanced”, using a rubric. This indicator measures the degree to which the “teacher often uses discussions and activities that encourage analysis and reasoning”. Additionally, indicator 21: “Creativity”. This indicator measures the degree to which the teacher “provides opportunities for students to be creative and/or generate their own ideas and products”.

CLASS
Under category “instructional support”, indicator 3: “Analysis and enquiry”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “facilitation of higher-order thinking”. This indicator refers to both “discussions and activities”.

FFT
Under category “instruction”, indicator 3b: “Using questioning and discussion techniques”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a
measure of the degree to which the teacher asks “questions with multiple correct answers or multiple approaches, even when there is a single correct response”.

SCOPE
Under category “language and literacy instruction”, indicator 13: “Writing instruction”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which "the teacher provides opportunities to produce text that expresses learners’ thoughts and reinforce their understanding”.

UTOP
Under category “classroom environment”, indicator 1.1: “Classroom engagement”. This indicator includes a measure of the degree to which “the teacher encouraged students to generate ideas, questions, conjectures, and/or propositions that reflected engagement or exploration with important mathematics and science concepts”. Rated on a scale from 1 to 5 using a rubric.

Coded as 0
We cannot find evidence of thinking tasks in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (29), Uganda (26), Uganda (27), Uganda (28), MQI, Stallings, SDI, TIPPS.

Sub-indicator 7.13 – Student thinking questions and/or answers
Definition: The students ask open ended questions and/or perform thinking tasks, such as, summarizing, comparing, contrasting, and/or analyzing content (TEACH).

Coded as 1
Brunei
Under category “teaching”, indicator 1.3: “Promote intellectual curiosity – + / –”.

Malawi
Indicator 11: “teaching for effective learning”. To score a 4 on this indicator, teachers need to “encourage students to think for themselves”. Although this indicator is rated from 1 to 4, it is done by a series of yes-no questions, therefore we code it as one.

Coded as 2
Tajikistan (29)
Under category “developing students personal competencies”, indicator II: “Critical thinking and problem solving”. Rated on a scale from 1 to 7 using a rubric. This indicator measures “opportunities for students to analyze a situation or a problem, create a plan, and implement a solution”. Additionally, under category “developing students personal competencies”, indicator IV: “Creativity and curiosity”. Also rated on a scale from 1 to 7 using a rubric.

FFT
Under category “instruction”, indicator 3b: “Using questioning and discussion techniques”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of the degree to which “students initiate higher-order questions”, “students extend the discussion, enriching it”, and “students invite comments from their classmates during a discussion and challenge one-another’s thinking”.

MQI

Under category “common core aligned practices”, indicator 2: “Student mathematical questioning and reasoning”. Rated on a scale of 4 from “not present” to “high” using a rubric. Additionally, under category “common core aligned practices”, indicator 4: “Task cognitive demand”. Rated on a scale of 4 from “not present” to “high” using a rubric. This indicator captures the degree of “student engagement in tasks in which they think deeply and reason about mathematics”.

Coded as 0

We cannot find evidence of student thinking questions and/or answers in the following tools: The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, CLASS, SCOPE, SDI, Stallings, TIPPS, UTOP.

Indicator 7.20 – Student focus

This is not a stand-alone indicator, meaning we only code this as one if one of the sub-indicators is also coded as a one.

Coded as 1

The Gambia

This tool received a one under sub-indicator 7.21, sub-indicator 7.23, and sub-indicator 7.24.

Kenya (20)

This tool received a one under sub-indicator 7.24.

Lebanon

This tool received a one under sub-indicator 7.21.

Malawi

This tool received a one under both sub-indicator 7.22 and sub-indicator 7.24.

Pakistan (58)

This tool received a one under both sub-indicator 7.23 and sub-indicator 7.24.

Tajikistan (30)

This tool received a one under sub-indicator 7.24.

Uganda (28)
This tool received a one under both sub-indicator 7.23 and sub-indicator 7.24.

Coded as 2
Guyana
This tool received a two under both sub-indicator 7.21 and sub-indicator 7.24.

Tajikistan (29)
This tool received a two under both sub-indicator 7.23 and sub-indicator 7.24.

Vietnam
This tool received a two under sub-indicator 7.22, sub-indicator 7.23, and sub-indicator 7.24.

CLASS
This tool received a two under sub-indicator 7.22, sub-indicator 7.23, and sub-indicator 7.24.

FFT
This tool received twos under all sub-indicators

MQI
This tool received a two under both sub-indicator 7.21 and sub-indicator 7.24.

SCOPE
This tool received a two under both sub-indicator 7.21 and sub-indicator 7.24.

TIPPS
This tool received a two under sub-indicator 7.21, sub-indicator 7.22, and sub-indicator 7.24.

UTOPOP
This tool received a two under sub-indicator 7.21, sub-indicator 7.23, and sub-indicator 7.24.

Coded as 0
We cannot find evidence of student focus in the following tools: Brunei, Liberia, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Philippines, Kenya (18), Kenya (19), Uganda (26), Uganda (27), SDI, Stallings.

Sub-indicator 7.21 – Active listening
Definition: The degree to which the teacher actively listens to student comments, responds appropriately, and encourages students’ ideas.

Coded as 1
The Gambia
Under category “student engagement and learning”, indicator 4: “Teacher encouraged students to ask questions and responded to them”. Also rated on a scale from 1 to 5 but with no rubric nor examples.

Lebanon
Under category “classroom management”, indicator 2: “Regular incorporation of students' ideas into the lesson – remarks”.

Coded as 2

Guyana
Under category “pedagogy”, indicator 3: “Learner ideas, experience and feedback are solicited/integrated into the lesson”. Rated on a scale from 0 to 4 using a rubric.

FFT
Under category “the classroom environment”, indicator 2a: “Creating an environment of respect and rapport”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “active listening”.

MQI
Under category “working with students and mathematics”, indicator 2: “Teacher uses student mathematical contributions”. Rated on a scale of 4 from “not present” to “high” using a rubric.

SCOPE
Under category “language and literacy instruction”, indicators 7 and 13: “Opportunities for oral language development”, and “writing instruction”. These indicators measure the degree to which the teacher encourages “a mutual exchange of ideas” and “learners’ thoughts and ideas” respectively. Both indicators are rated on a scale from 1 to 5 using a rubric.

TIPPS
Indicator 12B: “Behavioural indications of positive environment between teacher and students among peers”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 12A, which is also rated on the same scale. This indicator measures “active listening”.

UTOP
Under category “classroom environment”, indicator 1.1: “Classroom engagement”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the degree to which the teacher “encouraged students to generate ideas”.

Coded as 0

We cannot find evidence of active listening in the following tools:
Brunei, Kenya (18), Kenya (19), Kenya (20), Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), CLASS, SDI, Stallings.
**Sub-indicator 7.22 – Student autonomy**

Definition: The degree to which students can exert choice, take leadership opportunities, and have responsibilities in the classroom.

**Coded as 1**

**Malawi**

Indicator 16: “School self-evaluation and improvement”. To score a 4 on this indicator, “students have opportunities to suggest possible improvements to the school during meetings and are informed about any action taken in response and the reasons.”. Although this indicator is rated from 1 to 4, it is done by a series of yes-no questions, therefore we code it as one. Additionally, indicator 11: “Teaching for effective learning”. To score a 4 on this indicator, teachers need to be “encourage students to think for themselves”. Although this indicator is rated from 1 to 4, it is done by a series of yes-no questions, therefore we code it as one.

**Coded as 2**

**Vietnam**

Indicators 10 and 13: “Teacher facilitation”, and “student leadership. Both indicators are rated on a scale of 3 from “basic” to “advanced”. Indicator 10 includes a measure of how “teacher-led vs. student-let” the class is.

**CLASS**

Under category “emotional support”, indicator 3: “Regard for student perspectives”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the degree of “support of [student] autonomy”.

**FFT**

Under category “the classroom environment”, indicator 2c: “Managing classroom procedures”. This indicator measures the degree to which “students [are] playing an important role in carrying out the routines”. Rated on a scale from 1 to 4 using a rubric.

**TIPPS**

Indicator 18B: “Teacher employs effective classroom management to create an environment that is conducive to learning”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 18A, which is also rated on the same scale. This indicator measures the degree to which the teacher allows “students to participate in classroom management by assigning responsibilities”.

**Coded as 0**

We cannot find evidence of student autonomy in the following tools:

Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), MQI, SCOPE, SDI, Stallings, UTOP.
**Sub-indicator 7.23 – Student ease in educational environment**
Definition: The degree to which the teacher allows students to share their ideas in class, ask for help, or take risks.

**Coded as 1**
The Gambia
Under category “student engagement and learning”, indicator 4: “Teacher encouraged students to ask questions and responded to them”. Also rated on a scale from 1 to 5 but with no rubric nor examples.

Pakistan (58)
Under category “the pupils”, indicator 5: “Were the students encouraged to ask questions? – yes, no, to some extent”.

Uganda (28)
Under category “teacher-student interactions and participation”, indicator 9: “Teacher encouraged the students to ask him/her questions”. Rated on a scale from 1 to 4 but with no rubric nor examples.

**Coded as 2**
Tajikistan (29)
Under category “developing students personal competencies”, indicator III: “Confidence and leadership”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the degree to which “students feel comfortable and at ease”.

CLASS
Under category “emotional support”, indicator 2: “Teacher sensitivity”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the degree of “student comfort” in “sharing their ideas” and “respond[ing] to questions”.

FFT
Under category “the classroom environment”, indicator 2a: “Creating an environment of respect and rapport”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of the degree to which students “are comfortable taking intellectual risks”.

UTOP
Under category “classroom environment”, indicator 1.6: “Classroom equity”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the degree to which “students were comfortable sharing their ideas”.

**Coded as 0**
We cannot find evidence of student ease in educational environment in the following tools: Brunei, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (30), Uganda (26), Uganda (27), MQI, SCOPE, SDI, Stallings, TIPPS.
**Sub-indicator 7.24 – Student engagement**

Definition: The degree to which students are actively engaged in class activities most of the time, as demonstrated by students paying attention, raising hands, answering questions, participating in group work, or the like.

**Coded as 1**

The Gambia

Under category “classroom management”, indicator 3: “Captured students’ attention and encouraged active participation”. Rated on a scale from 1 to 5, but with no rubric nor examples. Additionally, under category “student engagement and learning”, indicator 1: “Students were paying attention and gave indications that they were learning”. Also rated on a scale from 1 to 5 but with no rubric nor examples.

Kenya (20)

“Participation of girls in the lesson – 81-100%, 61-80%, 41-60%, 21-40%, 1-20%”, and “participation of boys in the lesson – 81-100%, 61-80%, 41-60%, 21-40%, 1-20%”. Additionally, “teacher uses paired or group work to encourage active participation of students – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Malawi

Indicator 1: “Learning in lessons”. To score a 2 or above on this indicator, students need to be “on task and respond actively in lessons”. Although this indicator is rated from 1 to 4, it is done by a series of yes-no questions, therefore we code it as one.

Pakistan (58)

Under category “the pupils”, indicator 4: “How many students put their hands up to answer? – most, half, few”.

Tajikistan (30)

Indicator 26: “During the lesson students are involved in active learning process – a) all the time, b) 75-99% of the time, c) 50-74% of the time, d) 25-49%, e) less than 25%, f) difficult to determine”.

Uganda (28)

Under category “teacher-student interactions and participation”, indicator 12: “Learners showed interest and actively participated during the lesson”. Rated on a scale from 1 to 4 but with no rubric or examples.

**Coded as 2**

Guyana

Under category “pedagogy”, indicator 6: “Learners are actively engaged in lesson”. Rated on a scale from 0 to 4 using a rubric.

Tajikistan (29)

Under category “Lesson strategies and delivery”, indicator I: “Effective facilitation”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the
degree to which “students are consistently interested and involved in activities and lessons”.

Vietnam

Indicator 9: “Maximising learning time”. Rated on a scale of 3 from “basic” to “advanced”.

CLASS

Under category “instructional support”, indicator 1: “Instructional learning formats”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “student engagement”.

FFT

Under category “instruction”, indicator 3b: “Using questioning and discussion techniques”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “student participation”. Additionally, under category “instruction”, indicator 3c: “Engaging students in learning”. Also rated on a scale from 1 to 4 using a rubric.

MQI

Under category “common core aligned student practices”, indicator 3: “Students communicate about the mathematics of the segment”. Rated on a scale of 4 from “not present” to “high”. This indicator is a measure of “student contributions”. Additionally, under category “whole lesson codes”, indicator 3: “Students are engaged”. Also rated on a scale of 4 from “not present” to “high”.

SCOPE

Under category “classroom structure”, indicator 3: “Participation of all learners”. Rated on a scale from 1 to 5 using a rubric.

TIPPS

Indicator 19B: “Students are engaged in classroom learning activities”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 19A, which is also rated on the same scale.

UTOP

Under category “classroom environment”, indicator 1.1: “Classroom engagement”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the degree to which there was “deep student engagement with the mathematics and/or scientific content”.

Coded as 0

We cannot find evidence of student engagement in the following tools:
Brunei, Kenya (18), Kenya (19), Lebanon, Liberia, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Philippines, Uganda (26), Uganda (27), SDI, Stallings.
**Indicator 8.10 – Positive social climate**
This is not a stand-alone indicator, meaning we only code this as one if one of the sub-indicators is also coded as a one.

*Coded as 1*
- Brunei
  This tool received a one under both sub-indicator 8.11 and sub-indicator 8.13.
- The Gambia
  This tool received a one under both sub-indicator 8.11, and sub-indicator 8.13.
- Kenya (20)
  This tool received a one under both sub-indicator 8.11 and sub-indicator 8.13.
- Lebanon
  This tool received a one under sub-indicator 8.11.
- Malawi
  This tool received a one under sub-indicator 8.11.
- Pakistan (58)
  This tool received a one under both sub-indicator 8.11, and sub-indicator 8.13.
- Tajikistan (30)
  This tool received a one under sub-indicator 8.11.
- Uganda (26)
  This tool received a one under sub-indicator 8.13.
- Uganda (27)
  This tool received a one under sub-indicator 8.13.
- Uganda (28)
  This tool received a one under sub-indicator 8.13.
- SDI
  This tool received a one under sub-indicator 8.13.

*Coded as 2*
- Guyana
  This tool received a two under sub-indicator 8.13.
- Philippines
  This tool received a two under both sub-indicator 8.11 and sub-indicator 8.13.
- Tajikistan (29)
  This tool received a two under all sub-indicators.
Vietnam
This tool received a two under both sub-indicator 8.11 and sub-indicator 8.12.

CLASS
This tool received a two under all sub-indicators.

FFT
This tool received a two under all sub-indicators.

MQI
This tool received a two under sub-indicator 8.11.

SCOPE
This tool received a two under sub-indicator 8.13.

TIPPS
This tool received a two under all sub-indicators.

UTOP
This tool received a two under sub-indicator 8.12.

Coded as 0
We cannot find evidence of positive social climate in the following tools:
Liberia, Kenya (18), Kenya (19), Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Stallings.

Sub-indicator 8.11 – Teacher awareness and responsiveness
Definition: The degree to which the teacher demonstrates awareness of student needs or problems, anticipates issues that students might have, offers assistance or support to individual students, is effective in addressing student problems, responds quickly to student needs, and adjusts lesson speed or wait time.

Coded as 1
Brunei
Under category “demonstrate high standards of personal and professional conduct”, indicator 6.2: “acknowledge the concerns, values and priorities of students’ families, cultures, and communities – + / –”.

The Gambia
Under category “classroom management”, indicator 8: “Moved around the classroom to check pupils’ work and supported those needing help”. Rated on a scale from 1 to 5 but with no rubric nor examples.

Kenya (20)
“Teacher moves around the classroom to interact with individual students – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Lebanon
Under category “classroom management”, indicator 4: “Purposeful movement around the classroom – remarks”.

Malawi
Indicator 22: “Care and welfare of the students”. To score a 4 on this indicator, “students have a trusted member of staff to whom they can go with problems and receive advice”. Although this indicator is rated from 1 to 4, it is done by a series of yes-no questions, therefore we code it as one.

Pakistan (58)
Under category “the teacher”, indicator 2: “Did the teacher really try to pay attention to all the pupils? – yes, no, to some extent”. Additionally, under category “the teacher”, indicator 5: “Was the teacher aware of the problems and difficulties faced by the pupils? – yes, no, to some extent”.

Tajikistan (30)
Indicator 23: “Teacher paid attention to … of students during the practicing of the learning skill – a) to all, b) to more than 50%, c) to less than 50%, d) to some”.

Coded as 2
Philippines
Indicator 16: “Plans and delivers teaching strategies that are responsive to the special education needs of learners in difficult circumstances, including: geographic isolation; chronic illness; displacement due to armed conflict, urban resettlement, or disasters; child abuse and child labour practices”. Rated on a scale from 1 to 9 using a rubric.

Tajikistan (29)
Under category “classroom climate”, indicator III: “Teacher sensitivity”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the degree to which the teacher is “aware of student needs”.

Vietnam
Indicator 11: “Teacher responsiveness”. Rated on a scale of 3 from “basic” to “advanced” using a rubric.

CLASS
Under category “emotional support”, indicator 2: “Teacher sensitivity”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of teacher “awareness”, and “responsiveness to academic and social/emotional needs and cues”.

FFT
Under category “the classroom environment”, indicator 2d: “Managing student behaviour”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “monitoring student behaviour”, and “teacher awareness of student conduct”.

Under category “working with students and mathematics”, indicator 2: “Teacher uses student mathematical contributions”. Rated on a scale of 4 from “not present” to “high”. This indicator measure the degree to which the “teacher “hears” what students are saying, mathematically, and responds appropriately during instruction”.

TIPPS
Indicator 13B: “Teacher is sensitive and responsive to students’ needs and learning”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 13A, which is also rated on the same scale.

We cannot find evidence of teacher awareness and responsiveness in the following tools: Kenya (18), Guyana, Kenya (19), Liberia, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Uganda (26), Uganda (27), Uganda (28), SCOPE, SDI, Stallings, UTOP.

Sub-indicator 8.12 – Student social competence
Definition: The degree to which students display self-regulation and interpersonal knowledge.

Tajikistan (29)
Under category “developing students personal competencies”, indicator I: “Communication”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the degree to which “students demonstrate respect and responsibility consistently when communicating with others and the teacher”.

Vietnam
Indicator 15: “Teamwork/cooperative learning”. Rated on a scale of 3 from “basic” to “advanced” using a rubric. This indicator includes measures of “Interaction among students”, “working together rather than working in parallel”, and “supporting one- another”.

CLASS
Under category “emotional support”, indicator 1: “Positive climate”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “positive peer interactions”.

FFT
Under category “the classroom environment”, indicator 2a: “Creating an environment of respect and rapport”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “student interaction with other students, including both words and actions”.

TIPPS
Indicator 12B: “There are behavioural indicators of positive environment between teacher and pupils and among peers”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 12A, which is also rated on the same scale.

UTOP

Under category “classroom environment”, indicator 1.2: “Classroom interactions”. Rated on a scale from 1 to 5 using a rubric. This indicator measures the degree to which “interactions reflected collegial working relationships among students”.

Coded as 0

We cannot find evidence of student social competence in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), MQI, SCOPE, SDI, Stallings.

Sub-indicator 8.13 – Positive student-teacher interaction
Definition: The degree to which the teacher and students are physically close to each other, engage in social and rapport conversation, display warm/supportive interactions, offer each other praise and encouragement, listen to each other, cooperate, use names and respectful language/tone, and smile or laugh. Degree to which the classroom is free of anger, harsh voices, physical aggression, teacher threats or physical control of students, disrespect, bullying, teasing, and sarcasm.

Coded as 1

Brunei

Under category “engage students in intellectually challenging learning”, indicator 1.5: “Promote the belief that the students can be successful – + / –”. Additionally, under category “suit work to students’ individual needs – + / –”, indicator 3.3: “respects students as individuals – + / –”. Additionally, under category “adopt and refine professional practice – + / –”, indicator 5.4a: and “promote mutual respect between teacher and students”.

The Gambia

Under category “classroom management”, indicators 5, 6, and 7: “Minimised distractions and responded well to disruptive behaviour”, “did not use corporal punishment or other forms of humiliating treatment”, and “used students’ names when asking questions”. All of these indicators are rated on a scale from 1 to 5 but with no rubric nor examples.

Kenya (20)

“Teacher displays positive tone by using encouragement rather than criticism – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”. “Teacher knows and uses student names – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”. “Teacher relate well to learners – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.
Pakistan (58)
Under category “the teacher”, indicator 1: “Did the teacher manage to create a positive relationship with the class? – yes, no, to some extent”.

Uganda (26)
Under category “good participant/trainer relationship”, indicators 1 through 4: “approachable”, “friendly”, “considerate”, “respect for participants’ views”. Rated on a scale of 3 from “well developed” to “needs development”, but with no rubric nor examples.

Uganda (27)
Under category “quality of teaching and learning process”, indicator 5: “the quality of teacher-learner interaction”. Rated on a scale from 1 to 4 but with no rubric nor examples.

Uganda (28)
Under category “teacher-student interactions and participation”, indicators 10 and 11: “Teacher demonstrated an open and respectful attitude towards learners”, and “teacher interacted with students, he/she made eye contact and knows some names”. Rated on a scale from 1 to 4 but with no rubric nor examples.

SDI

Coded as 2
Guyana
Under category “classroom and student management”, indicator 3: “Classroom interactions reflect interpersonal respect and learner support”. Rated on a scale from 0 to 4 using a rubric.

Philippines
Indicator 10: “Maintains supportive learning environments that nurture and inspire learners to participate, cooperate, and collaborate in continued learning”. Rated on a scale from 1 to 9 using a rubric.

Tajikistan (29)
Under category “classroom climate”, indicator 1: “Positive climate”. Rated on a scale from 1 to 7 using a rubric.

CLASS
Under category “emotional support”, indicator 1: “Positive climate”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of teacher “respect”, and “positive communications”.
FFT
Under category “the classroom environment”, indicator 2a: “Creating an environment of respect and rapport”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “teacher interaction with students, including both words and actions”.

SCOPE
Under category “classroom culture”, indicator 1: “Positive learning environment”. Rated on a scale from 1 to 5 using a rubric.

TIPPS
Indicator 12B: “There are behavioural indicators of positive environment between teacher and pupils and among peers”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 12A, which is also rated on the same scale.

Coded as 0
We cannot find evidence of positive student-teacher interaction in the following tools:
Kenya (18), Kenya (19), Liberia, Lebanon, Malawi, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (30), Vietnam, MQI, Stallings, UTOP.

Indicator 8.20 – Equality and Inclusiveness
Definition: The degree to which a teacher purposefully treats all students in the same manner, and absolves from favouring some students over others.

Coded as 1
The Gambia
Under category “classroom management”, indicator 9: “Treated all students equally”. Rated on a scale from 1 to 5 but with no rubric nor examples.
Kenya
This tool received a one under both sub-indicator 8.21 and sub-indicator 8.23.
Liberia
This tool received a one under sub-indicator 8.21.
Malawi
This tool received a one under both sub-indicator 8.21 and sub-indicator 8.23.
Uganda
This tool received a one under sub-indicator 8.23.

Coded as 2
Philippines
Indicator 20: “Maintains learning environments that promote fairness, respect and care, to encourage learning”. Rated on a scale from 1 to 9 using a rubric.
FFT
This tool received a two under sub-indicator 8.23.

SCOPE
This tool received a two under sub-indicator 8.21, sub-indicator 8.22, and sub-indicator 8.23.

TIPPS
This tool received a two under both sub-indicator 8.21 and sub-indicator 8.23.

UTOP
This tool received a two under sub-indicator 8.21, sub-indicator 8.22, and sub-indicator 8.23.

Coded as 0
We cannot find evidence of equality and inclusiveness in the following tools: Brunei, Guyana, Kenya (18), Kenya (19), Lebanon, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (28), Vietnam, CLASS, MQI, SDI, Stallings.

Sub-indicator 8.21 – Gender inclusiveness
Definition: The degree to which the teacher fosters a classroom environment that is unbiased toward gender.

Coded as 1
Kenya (20)
“Participation of girls in the lesson – 81-100%, 61-80%, 41-60%, 21-40%, 1-20%”, and “participation of boys in the lesson – 81-100%, 61-80%, 41-60%, 21-40%, 1-20%”. Additionally, “teacher encourages equal participation of male and female students – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”. “Teacher provides equal amount of feedback to both male and female students – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Liberia
Indicator 5.9: “Is the teacher giving both boy and girl students an opportunity to speak and ask questions? – yes, no”

Malawi
Indicator 3: “Attainment across the school”. To score a 4 on this indicator, the school must ensure “girls and boys attain equally well”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, therefore we code it as one.

Coded as 2
Philippines
Indicator 13: “Uses differentiated, developmentally appropriate learning experiences to address learners’ gender, needs, strengths, interests, and experiences”. Rated on a scale from 1 to 9 using a rubric.

SCOPE
Under category “classroom structure”, indicator 3: “Participation of all learners”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which “individual differences (e.g. cultural, racial, disability, language, gender, and ability level) are valued and addressed in multiple ways throughout the classroom to advance learners’ knowledge and development”.

TIPPS
Indicator 16B: “Teacher actively employs gender responsive strategies”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 16A, which is also rated on the same scale.

UTOOP
Under category “classroom environment”, indicator 1.6: “Classroom equity”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which “the classroom environment was unbiased related to race, ethnicity, religion, gender, sexual orientation, physical abilities, English language learners, and students with learning differences”.

Coded as 0
We cannot find evidence of gender inclusiveness in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Lebanon, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (26), Uganda (27), Uganda (28), Vietnam, CLASS, FFT, MQI, SDI, Stallings.

Sub-indicator 8.22 – Racial inclusiveness
Definition: The degree to which the teacher fosters a classroom environment that is unbiased toward race.

Coded as 1

Coded as 2
Philippines
Indicator 17: “Adapts and uses culturally appropriate learning strategies to address the needs of learners from indigenous groups”. Rated on a scale from 1 to 9 using a rubric.

SCOPE
Under category “classroom structure”, indicator 3: “Participation of all learners”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which “individual differences (e.g. cultural, racial, disability, language, gender, and ability level) are valued and addressed in multiple ways throughout the classroom to advance learners’ knowledge and development”.

Coded as 1
UTOP

Under category “classroom environment”, indicator 1.6: “Classroom equity”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which “the classroom environment was unbiased related to race, ethnicity, religion, gender, sexual orientation, physical abilities, English language learners, and students with learning differences”.

Coded as 0

We cannot find evidence of racial inclusiveness in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, CLASS, FFT, MQI, SDI, Stallings, TIPPS.

Sub-indicator 8.23 – Disability inclusiveness

Definition: The degree to which the teacher fosters a classroom environment that is unbiased toward disabled students, or students with learning difficulties.

Coded as 1

Kenya (20)
“Students with special educational needs identified and included in teacher questioning – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”.

Malawi

Indicator 3: “Attainment across the school”. To score a level 2 or above, the school must ensure “students’ with special educational needs are making progress in line with their previous attainment”. Although this indicator is rated from 1-4, it is done by a series of yes-no questions, therefore we code it as one.

Uganda (27)
Under category “quality of teacher’s planning”, indicator 6: “The extent to which lessons reflect the diverse special needs of learners”. Additionally, under category “quality of teaching and learning process”, indicator 9: “Activities differentiated to take account of special needs”. Both indicators are rated on a scale from 1 to 4 but with no rubric nor examples.

Coded as 2

Philippines

Indicator 15: “Designs, adapts, and implements teaching strategies that are responsive to learners with special education needs that include giftedness, talents and disabilities”. Rated on a scale from 1 to 9 using a rubric.

FFT
Under category “planning and preparation”, indicator 1b: “Demonstrating knowledge of students”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “knowledge of students’ special needs”.

**SCOPE**
Under category “classroom structure”, indicator 3: “Participation of all learners”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which “individual differences (e.g. cultural, racial, disability, language, gender, and ability level) are valued and addressed in multiple ways throughout the classroom to advance learners’ knowledge and development”.

**TIPPS**
Indicator 17B: “Teacher actively employs responsive strategies for diverse learners”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 17A, which is also rated on the same scale.

**UTO**
Under category “classroom environment”, indicator 1.6: “Classroom equity”. Rated on a scale from 1 to 5 using a rubric. This indicator includes a measure of the degree to which “the classroom environment was unbiased related to race, ethnicity, religion, gender, sexual orientation, physical abilities, English language learners, and students with learning differences”.

**Coded as 0**
We cannot find evidence of disability inclusiveness in the following tools: Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Lebanon, Liberia, Pakistan (57), Pakistan (58), Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (28), Vietnam, CLASS, MQI, SDI, Stallings.

**Indicator 8.30 – Motivating the classroom**
This is not a stand-alone indicator, meaning we only code this as one if one of the sub-indicators is also coded as a one.

**Coded as 1**
Brunei
This tool received a one under both sub-indicator 8.31 and sub-indicator 8.34.

Kenya (20)
This tool received a one under sub-indicator 8.34.

Lebanon
This tool received a one under sub-indicator 8.34.

**Coded as 2**
Tajikistan (29)
This tool received a two under both sub-indicator 8.32 and sub-indicator 8.34.

CLASS
This tool received a two under both sub-indicator 8.32 and sub-indicator 8.34.

FFT
This tool received a two under both sub-indicator 8.31 and sub-indicator 8.32.

Coded as 0
We cannot find evidence of motivating the classroom in the following tools:
The Gambia, Guyana, Kenya (18), Kenya (19), Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, MQI, SCOPE, SDI, Stallings, TIPPS, UTOP.

Sub-indicator 8.31 – Communicating high expectations
Definition: The degree to which the teacher imparts high expectations for student learning or demonstrates high-quality work.

Coded as 1
Brunei
Under category “engage students in intellectually challenging learning”, indicator 1.2: “have high expectations – + / – ”.

Coded as 2
FFT
Under category “the classroom environment”, indicator 2b: “Establishing a culture for learning”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “high expectations for expression and work products”.

Coded as 0
We cannot find evidence of communicating high expectations in the following tools:
The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, CLASS, MQI, SCOPE, SDI, Stallings, TIPPS, UTOP.

Sub-indicator 8.32 – Recognition of effort
Definition: The degree to which teacher recognizes a student's efforts to help others, assist in the classroom, and go above and beyond in their schoolwork.

Coded as 1
None

Coded as 2
Tajikistan (29)
Under category “assessment and evaluation”, indicator III: “Encouragement and affirmation”. Rated on a scale from 1 to 7 using a rubric. This indicator includes
measure of the degree to which the teacher “offers encouragement of students' efforts”.

CLASS
Under category “instructional support”, indicator 4: “Quality of feedback”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “recognition and affirmation of effort”.

FFT
Under category “the classroom environment”, indicator 2b: “Establishing a culture for learning”. Rated on a scale from 1 to 4 using a rubric. This indicator includes a measure of “expectation and recognition of effort and persistence on the part of students”.

Coded as 0
We cannot find evidence of recognition of effort in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, MQI, SCOPE, SDI, Stallings, TIPPS, UTOP.

Sub-indicator 8.33 – Student goal setting
Definition: The degree to which the teacher encourages goal setting (TEACH).

Coded as 1
None

Coded as 2
None

Coded as 0
We cannot find evidence of student goal setting in the following tools:
Brunei, The Gambia, Guyana, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, CLASS, FFT, MQI, SCOPE, SDI, Stallings, TIPPS, UTOP.

Sub-indicator 8.34 – Positive language and morale
Definition: The degree to which the teacher uses positive language with students and fosters high morale. This indicator refers specifically to positive verbal communication, as opposed to indicator 8.13, positive teacher-student interaction which includes positive environment and non-verbal communication.

Coded as 1
Brunei
Under category “engage students in intellectually challenging learning”, indicator 1.5: “promote the belief that students can be successful – + / –”.
Kenya (20)
“Teacher displays positive tone by using encouragement rather than criticism – behaviour never observed, behaviour rarely observed, behaviour occasionally observed, behaviour consistently observed”

Lebanon
Under category “classroom management”, indicator 5: “Positive reinforcement and feedback – remarks”.

Coded as 2

Tajikistan (29)
Under category “classroom climate”, indicator 1: “Positive climate”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “positive verbal and physical communication”, and the degree to which the teacher uses “positive comments”.

CLASS
Under category “emotional support”, indicator 1: “Positive climate”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “positive communications”.

Coded as 0

We cannot find evidence of positive language and morale in the following tools:
The Gambia, Guyana, Kenya (18), Kenya (19), Liberia, Malawi, Pakistan (57), Pakistan (58), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (30), Uganda (26), Uganda (27), Uganda (28), Vietnam, FFT, MQI, SCOPE, SDI, Stallings, TIPPS, UTOP.

Indicator 8.40 – Family engagement
Definition: The degree to which the teacher involves family in student learning, when appropriate.

Coded as 1

Brunei
Under category “demonstrate high standards of personal and professional conduct”, indicator 6.2: “Acknowledge the concerns, values and priorities of students’ families, cultures and communities – + / –”

Malawi
Indicator 16: “School self-evaluation and improvement”. To score a 3 or above on this indicator, the “the SMC/governing body encourages parents and the community to be involved in the school, by organising activities and reporting on school event and achievements”. Although this indicator is rated from 1 to 4, it is done by a series of yes-no questions, therefore we code it as one.

Pakistan (58)
Under category “teacher appraisal”, indicator 11: “Relationship with parents”. Rated on a scale of 3 from “acceptable or competent” to “unacceptable”.

Uganda (27)
Under category “assessment and record keeping”, indicator 6: “Learners’ achievement is communicated to parents”. Rated on a scale from 1 to 4 but with no examples nor a rubric.

Coded as 2
Vietnam
Indicator 7: “Evaluation and feedback”. Rated on a scale of 3 from “basic” to “high”. This indicator includes a measure of the degree to which “the teacher informs each parent about student’s progress”.

FFT
Under category “professional responsibilities”, indicator 4c: “Communicating with families”. Rated on a scale from 1 to 4 using a rubric.

Coded as 0
We cannot find evidence of family engagement in the following tools:
The Gambia, Kenya (18), Kenya (19), Kenya (20), Lebanon, Liberia, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Philippines, Tajikistan (29), Tajikistan (30), Uganda (26), Uganda (28), CLASS, MQI, SCOPE, SDI, Stallings, UTOP.

**Indicator 9.10 – Professionalism and Respect**
Definition: The degree to which the teacher treats students and colleagues with respect and courtesy and demonstrates rapport and mutual trust with colleagues and parents.

Coded as 1
Brunei
Under category “demonstrate high standards of personal and professional conduct”, indicators 6.3, 6.4, and 6.5: “Develop professional relationships (colleagues, parents and communities) – + / –”, “maintain high standards of attendance and punctuality – + / –”, and “keep well informed regarding official policies, regulations and other requirements – + / –”.

Kenya (20)
Under category “professionalism of the teacher”, indicators 5.1, and 5.2: “Teacher’s professional appearance – unsatisfactory, satisfactory, good, very good”, and “teacher’s punctuality and time keeping – unsatisfactory, satisfactory, good, very good”.

Pakistan (58)
Under category “teacher appraisal”, indicator 10: “Professional responsibility”. Rated on a scale of 3 from “acceptable or competent” to “unacceptable”.


Uganda (26)
Under category “personality”, indicator 1: “Respect for self and others”. Rated on a scale of 3 from “needs development” to “well developed”, but with no rubric nor examples.

Uganda (28)
Under category “teacher-student interactions and participation”, indicator 10: “Teacher demonstrated an open and respectful attitude towards learners”. Rated on a scale from 1 to 4 but with no rubric nor examples.

Coded as 2
Guyana
Under category “classroom and student management”, indicator 3: “classroom interactions reflect interpersonal respect and learner support”. Rated on a scale from 0 to 4 using a rubric.

Philippines
Indicator 20: “Maintains learning environments that promote fairness, respect and care, to encourage learning”. Rated on a scale from 1 to 9 using a rubric.

Tajikistan (29)
Under category “classroom climate”, indicator I: “Positive climate”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of the degree to which “the teacher and students consistently demonstrate respect for one another”.

CLASS
Under category “emotional support”, indicator 1: “Positive climate”. Rated on a scale from 1 to 7 using a rubric. This indicator includes a measure of “respect”.

FFT
Under category “professional responsibilities”, indicator 4f: “Showing professionalism”. Rated on a scale from 1 to 4 using a rubric.

TIPPS
Indicator 14A: “Behavioral indications of negative environment between the teacher and students and amongst peers”. Rated using a rubric, from “somewhat accurate” to “very accurate”, against its converse, indicator 14B, which is also rated on the same scale. This indicator includes a measure of “respect for students”.

Coded as 0
We cannot find evidence of professionalism and respect in the following tools: The Gambia, Guyana, Kenya (18), Kenya (19), Lebanon, Liberia, Malawi, Pakistan (57), Papua New Guinea (10), Papua New Guinea (11), Tajikistan (30), Uganda (27), Vietnam, MQI, SCOPE, SDI, Stalling, UTOP.