Classroom Observation Toolkit for Early Grade Reading Improvement

A Global Reading Network Resource

This resource was made possible by the support of the American people through the United States Agency for International Development (USAID). This resource was prepared for USAID’s Building Evidence and Supporting Innovation to Improve Primary Grade Reading Assistance for the Office of Education (E3/ED), University Research Co., LLC, Contract No. AID-OAA-M-14-00001, MOBIS#: GS-10F-0182T.
Cover photo credit:
Kelley Lynch, Global Partnership for Education
Classroom Observation Toolkit for Early Grade Reading Improvement

A Global Reading Network Resource

Ashley Clayton Hertz
Emily Kochetkova
Alison Pflepsen

November 2019
Except as otherwise noted, *Classroom Observation Toolkit for Early Grade Reading Improvement* is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-sa/4.0/.

Under the Creative Commons Attribution-NonCommercial-ShareAlike license, you are free to share and adapt the material in any medium or format under the following conditions:

**Attribution:** You must give appropriate credit, provide a link to the licensed material and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. A suggested citation appears at the bottom of this page, and please also note: This document was developed by Reading within Reach, through the support of the U.S. Agency for International Development (USAID) and licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

**Non-Commercial:** You may not use the material for commercial purposes.

**Translations:** If you create a translation of this work, please add the following disclaimer along with the attribution: This translation was not created by USAID and should not be considered an official USAID translation. USAID shall not be liable for any content or error in this translation.

**Adaptations:** If you create an adaptation of this work, please add the following disclaimer along with the attribution: This is an adaptation of an original work by USAID. Views and opinions expressed in the adaptation are the sole responsibility of the author or authors of the adaptation and are not endorsed by USAID.

**Third-Party Content:** USAID does not necessarily own each component of the content contained within the work. USAID therefore does not warrant that the use of any third party–owned individual component or part contained in the work will not infringe on the rights of those third parties. The risk of claims resulting from such infringement rests solely with you. If you wish to reuse a component of the work, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright owner. Examples of components can include, but are not limited to, tables, figures or images.

**Disclaimer:** The views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development.

**Suggested citation:** Hertz, Ashley C., Emily Kochetkova, and Alison Pflepsen. Classroom Observation Toolkit for Early Grade Reading Improvement. A Global Reading Network Resource. Prepared by University Research Co., LLC. (URC) under the Reading within REACH initiative for USAID's Building Evidence and Supporting Innovation to Improve Primary Grade Assistance for the Office of Education (E3/ED). www.globalreadingnetwork.net.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLES AND FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>ACRONYMS</td>
<td>x</td>
</tr>
<tr>
<td>GLOSSARY OF TERMS</td>
<td>xi</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>xiii</td>
</tr>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>1</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>UNDERSTANDING CLASSROOM OBSERVATION</td>
<td>4</td>
</tr>
<tr>
<td>Defining Classroom Observation</td>
<td>4</td>
</tr>
<tr>
<td>Importance of Classroom Observation</td>
<td>4</td>
</tr>
<tr>
<td>Purpose and Use of Classroom Observation</td>
<td>5</td>
</tr>
<tr>
<td>Identify Teacher Professional Development Needs</td>
<td>5</td>
</tr>
<tr>
<td>Support Individual Teacher Professional Growth and Learning</td>
<td>6</td>
</tr>
<tr>
<td>Monitor Program Implementation</td>
<td>6</td>
</tr>
<tr>
<td>Evaluate the Impact of Professional Development</td>
<td>6</td>
</tr>
<tr>
<td>Measure and Improve the Quality of Teaching Instruction at a System Level</td>
<td>7</td>
</tr>
<tr>
<td>Generate New Knowledge About Teaching-Learning Processes</td>
<td>7</td>
</tr>
<tr>
<td>Evaluate Individual Teacher Performance for Accountability</td>
<td>7</td>
</tr>
<tr>
<td>Instrument Features and Content</td>
<td>8</td>
</tr>
<tr>
<td>Validity and Reliability</td>
<td>8</td>
</tr>
<tr>
<td>Item and Response Formats</td>
<td>8</td>
</tr>
<tr>
<td>Focus of Instrument</td>
<td>9</td>
</tr>
<tr>
<td>Instructional Categories and Observable Behaviors</td>
<td>11</td>
</tr>
<tr>
<td>Feasibility and Usability</td>
<td>12</td>
</tr>
<tr>
<td>EGR PROGRAM EXPERIENCES AND GUIDANCE</td>
<td>14</td>
</tr>
<tr>
<td>Purpose of Classroom Observations in EGR Programs</td>
<td>14</td>
</tr>
<tr>
<td>Experiences from EGR Programs</td>
<td>15</td>
</tr>
<tr>
<td>Guidance on Instrument Purpose and Use</td>
<td>17</td>
</tr>
<tr>
<td>Observation Instrument Features and Content</td>
<td>19</td>
</tr>
<tr>
<td>Experiences from EGR Programs</td>
<td>19</td>
</tr>
<tr>
<td>Guidance on Instrument Features and Content</td>
<td>26</td>
</tr>
<tr>
<td>Instrument Development and Adaptation</td>
<td>30</td>
</tr>
<tr>
<td>Experiences from EGR Programs</td>
<td>30</td>
</tr>
<tr>
<td>Guidance on Instrument Development and Adaptation</td>
<td>31</td>
</tr>
<tr>
<td>Preparation of Observers on Instrument Use</td>
<td>35</td>
</tr>
<tr>
<td>Experiences from EGR Programs</td>
<td>35</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Guidance on Observer Preparation</td>
<td>37</td>
</tr>
<tr>
<td>Instrument Administration</td>
<td>40</td>
</tr>
<tr>
<td>Experiences from EGR Programs</td>
<td>40</td>
</tr>
<tr>
<td>Guidance on Instrument Administration</td>
<td>42</td>
</tr>
<tr>
<td>Dissemination and Use of Results</td>
<td>44</td>
</tr>
<tr>
<td>Experiences from EGR Programs</td>
<td>44</td>
</tr>
<tr>
<td>Guidance on Dissemination and Use</td>
<td>45</td>
</tr>
<tr>
<td>Planning for Classroom Observation: Additional Considerations</td>
<td>46</td>
</tr>
<tr>
<td>SUMMARY AND FUTURE DIRECTIONS</td>
<td>49</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>51</td>
</tr>
<tr>
<td>ANNEXES</td>
<td>55</td>
</tr>
<tr>
<td>Annex A. Summary of Guidance for Instrument Development and Use</td>
<td>55</td>
</tr>
<tr>
<td>Annex B. Identifying Teacher Professional Development Needs</td>
<td>57</td>
</tr>
<tr>
<td>Annex C. Supporting Individual Teacher Growth</td>
<td>61</td>
</tr>
<tr>
<td>Annex D. Monitoring Teachers’ Adherence to EGR Program Approach</td>
<td>67</td>
</tr>
<tr>
<td>Annex E. Measuring Changes in Teachers’ Use of Instructional Time</td>
<td>71</td>
</tr>
<tr>
<td>Annex F. Simple Observation Checklist</td>
<td>77</td>
</tr>
</tbody>
</table>
# Tables and Figures

## Tables

Table 1. Instructional categories and examples of observable behaviors ................................................................. 13
Table 2. Classroom observation purposes common in EGR initiatives ........................................................................ 18
Table 3. Focus of observation instruments in EGR initiatives, by purpose .................................................................. 28
Table 4. Recommended item and response formats for EGR observation instruments, by purpose ............................... 29
Table 5. Stakeholders and experts to include in classroom observation instrument development, by purpose ........... 32
Table 6. Observer profiles and skills required, by purpose .......................................................................................... 39
Table 7. Frequency of classroom observation, by purpose .......................................................................................... 43

## Figures

Figure 1. Binary checklist with space for coaches to record comments and evidence .................................................. 20
Figure 2. Binary checklist for monitoring adherence .................................................................................................. 20
Figure 3. Checklist with three different options .......................................................................................................... 20
Figure 4. Time interval sampling format in evaluations ............................................................................................... 21
Figure 5. Analysis section of a coaching instrument ................................................................................................... 22
Figure 6. Automatically generated feedback in Tangerine:Tutor .................................................................................. 23
Figure 7. Dashboard from Unlock Literacy MEQA Program .......................................................................................... 42
## Acronyms

<table>
<thead>
<tr>
<th>DEC</th>
<th>Development Experience Clearinghouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDC</td>
<td>Education Development Center, Inc.</td>
</tr>
<tr>
<td>EGR</td>
<td>early grade reading</td>
</tr>
<tr>
<td>EGRA</td>
<td>Early Grade Reading Assessment</td>
</tr>
<tr>
<td>FHI 360</td>
<td>Family Health International 360</td>
</tr>
<tr>
<td>FOI</td>
<td>fidelity of implementation</td>
</tr>
<tr>
<td>GPS</td>
<td>global positioning system</td>
</tr>
<tr>
<td>GRN</td>
<td>Global Reading Network</td>
</tr>
<tr>
<td>IRC</td>
<td>International Rescue Committee</td>
</tr>
<tr>
<td>IRR</td>
<td>inter-rater reliability</td>
</tr>
<tr>
<td>LMIC</td>
<td>low- and middle-income country</td>
</tr>
<tr>
<td>LPWT</td>
<td>Literacy Practice Walk Through (Ghana)</td>
</tr>
<tr>
<td>M&amp;E</td>
<td>monitoring and evaluation</td>
</tr>
<tr>
<td>MEQA</td>
<td>Measuring Evidence of Quality Achieved</td>
</tr>
<tr>
<td></td>
<td>(Cambodia, Nepal, and Bangladesh)</td>
</tr>
<tr>
<td>MERIT</td>
<td>Malawi Early Grade Reading Improvement Activity</td>
</tr>
<tr>
<td>MOE</td>
<td>Ministry of Education</td>
</tr>
<tr>
<td>NEI</td>
<td>Northern Education Initiative (Nigeria)</td>
</tr>
<tr>
<td>PRP</td>
<td>Pakistan Reading Program</td>
</tr>
<tr>
<td>RARA</td>
<td>Reading and Access Research Activity (Nigeria)</td>
</tr>
<tr>
<td>REACH</td>
<td>Reading within Reach</td>
</tr>
<tr>
<td>REACH</td>
<td>Results in Education and Child Health</td>
</tr>
<tr>
<td>SHRP</td>
<td>School Health and Reading Program (Uganda)</td>
</tr>
<tr>
<td>SIRA</td>
<td>Selective Integrated Reading Activity (Mali)</td>
</tr>
<tr>
<td>URC</td>
<td>University Research Co., LLC</td>
</tr>
<tr>
<td>USA</td>
<td>United States of America</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
<tr>
<td>WASH</td>
<td>water, sanitation and hygiene</td>
</tr>
</tbody>
</table>
Glossary of Terms

**Binary checklists** are a type of response format on a classroom observation instrument. Binary checklists are simple checklists with questions or statements of items or behaviors that an individual intends to observe. The binary response format provides the individual with a choice of two responses, most commonly yes or no. However, some instruments use true or false as response options. Classroom observation is the process of an individual observing a teacher facilitating a lesson either in a classroom or another type of learning environment.

**Classroom observation** instrument refers to a document or tool that a person uses to record what he or she sees when observing a teacher or class. The instrument may focus on teacher behavior, student behavior, resource use and/or the classroom environment itself.

**Early Grade Reading Assessment (EGRA)** is a specific assessment to measure children’s foundational reading skills. EGRA is designed to assess the reading abilities of students in early primary grades (e.g., 1-3) to inform the design of reading interventions, as well as to monitor and evaluate efforts to improve reading outcomes.

**Equitable and inclusive** instruction includes behaviors or activities teachers use in the classroom to help ensure that all students have equal opportunities to participate and learn in each lesson, including those students who are at risk of marginalization due to gender, ethnicity or disability.

**Feasibility** refers to the potential of a classroom observation instrument to be used appropriately and effectively given the purpose, scope of the instrument, context, observer capacity, and time and period in which the observation will be conducted.

**Fidelity of implementation (FOI)** refers to a program’s adherence to its design and plans during implementation. When monitoring or measuring fidelity of implementation, programs can assess the adherence to design as well as the quality and frequency of activities, or the duration of an individual’s or group’s exposure to specific strategies.

**High-inference** refers to an item on a classroom observation instrument that requires the observer to make an evaluative judgment about certain behaviors or processes in the classroom. For example, an item such as “The teacher taught letter sounds effectively” requires the individual to do more than note whether the teacher taught letter sounds. Instead, the observer must make a judgment regarding the effectiveness of how the teacher taught letter sounds.

**Inter-rater reliability (IRR)**, in the context of classroom observation, refers to the degree or extent to which a group of observers (or “raters”) scores a given item the same. IRR assessment typically occurs during classroom observation training for some instruments. Training that results in high IRR scores is more likely to lead to high quality and reliable data that are accurate and consistent across raters.

**Item and response formats** refer to the type of question and answer format of a classroom observation instrument. Types of item and response formats include binary checklists, rating scales and time-sampling.

**Low-inference** refers to an observation item on a classroom observation instrument that does not require the individual to make an evaluative judgment. Low-inference statements are factual and based strictly on classroom behavior or processes that the observer can see. For example, “The teacher provides opportunity for students to work in groups.

**Rating scale** is a type of response format on a classroom observation instrument that allows the observer to rate the frequency or quality of specific items observed in the classroom. This requires the observer to make a judgment using specified criteria on a scale. Rating scales can use numerical numbers for scoring responses (1=never; 2=sometimes or 3=always) or qualitative responses (such as never; rarely; frequently and usually).
Protocols are the suggested steps or procedure observers can follow when conducting classroom observations. Protocols can include information about what to do before, during, and after a classroom observation and how to score the instrument. Some protocols might provide guidance on how to give feedback. Protocols for observation visits usually come in a documented form and are included with a classroom observation instrument, either on the instrument itself or as part of a separate manual.

Reliability refers to whether the instrument consistently measures what it intends to measure. To be considered reliable, the information collected on an observation instrument during a single observation should be consistent both in an individual’s scoring as well as across observers.

Safe learning environments refers to formal and non-formal learning environments and surrounding areas that are free from risks that may cause harm to students or personnel in the schools. Examples of hazards may include: natural hazards and disasters, health emergencies, school-related gender-based violence, gang activity, bullying, and individual or armed group attacks.

Social and emotional learning (SEL) is the process of an individual acquiring and applying the social, emotional and behavioral skills needed for optimal health and well-being. SEL interventions in formal and non-formal education programs support teachers to enhance the emotional health and well-being of children and youth, thus improving their opportunity to learn and succeed.

Structured observation instruments provide specific information on what an individual must observe and a specific format in which information is recorded. This contrasts with unstructured instruments (see below).

Structured pedagogy is a framework of specific teacher-led instructional principles designed to improve student achievement. Principles of structured pedagogy include: maximized use of instructional time, systematic and explicit instruction, instructional routines, scaffolding, assessment-informed decision-making, and social and emotional learning.

Time-interval sampling is a type of response format on a classroom observation instrument that allows an individual to record specific behaviors or processes and the frequency of those behaviors in a specified time interval.

Unstructured observation instruments do not have a specific behavior focus or item and response format. The instrument has an open-ended format that allows observers to collect descriptive, qualitative information about what they see during a lesson observation.

Usability refers to both the degree of “user-friendliness” of the classroom observation instrument (structure, format, and font) as well as the processes involved for data management and analysis. For example, an instrument that has small font and large amounts of text may be challenging for the observer to read. Some instruments might have user-friendly formats, but data management processes might be too difficult for easy data entry.

Validity is the extent to which a classroom observation instrument measures what it aims to measure. A classroom observation instrument that is valid captures instructional practices and behaviors that lead to improved student achievement.
The Classroom Observation Toolkit for Early Grade Reading Improvement was prepared by Reading within REACH (REACH), a five-year initiative funded by the U.S. Agency for International Development (USAID) and implemented by University Research Co., LLC (URC). The purpose of REACH is to support those designing and implementing early grade reading (EGR) initiatives in low- and middle-income contexts by providing resources and professional development opportunities; supporting innovations in early grade reading programming; and supporting the Global Reading Network, a community of practice that brings together practitioners, government and nongovernmental organizations, civil society groups and other stakeholders.

This Classroom Observation Toolkit was written by Ashley Clayton Hertz, technical consultant to REACH. Its development was supported by Emily Kochetkova, REACH technical consultant, and Alison Pflepsen, REACH Reading Program Specialist. The team would like to express its gratitude to the following individuals who reviewed an initial draft of the toolkit and provided feedback: Leslie Rosales Flores (Juarez & Associates), Julia Frazier (FHI 360), Cristine Smith (University of Massachusetts Amherst), and Kristina Solum (School-to-School International). Amy Pallangyo, REACH Technical Advisor, also provided helpful feedback and assistance. John Micklos edited the resource.

The Toolkit is informed by the rich experiences and lessons learned to date in EGR programming. This publication would not be possible without the valuable inputs and contributions from the following organizations: Columbia University, Creative Associates International, Education Development Center (EDC), EnCompass, FHI 360, International Rescue Committee (IRC), Juarez & Associates, Mathematica, Room to Read, RTI International, Social Impact, South Africa Department of Basic Education, World Bank and World Vision. In addition to sharing their experiences, many also shared resources, instruments and “best practices” from their own work. REACH expresses its deepest appreciation for their time and willingness to share resources with the broader community of practice.
Executive Summary

Classroom observation is the process of an individual observing a teacher facilitating a lesson either in a classroom or another type of learning environment. The overall goal of any classroom observation is to improve the quality of teaching and learning.

EGR programs primarily use classroom observations for the following purposes: to support teacher professional growth and learning through feedback; to monitor program implementation; and to evaluate the impact of programs. A review of practice to date in USAID-supported early grade reading improvement programs reveals a need for more purpose-oriented observations and higher quality instruments to help improve teacher performance in the classroom.

The purpose of this Classroom Observation Toolkit for Early Grade Reading Improvement is to provide guidance to USAID staff, USAID partners and governments partnering with USAID on how to improve the design, implementation and use of classroom observation instruments for the three purposes mentioned above. The Toolkit summarizes key principles related to classroom observation and provides guidance for developing high-quality observation instruments for use in primary grade reading interventions. This guidance is informed by a review of relevant literature and evaluation reports, responses from a survey on EGR classroom observation experience, and a sample of observation instruments from EGR programs globally. To assist governments and programs in developing high-quality observation instruments, the Toolkit also includes observation templates that can be adapted for different purposes and contexts.

Key recommendations on how EGR programs can design high-quality instruments and effectively apply the information gathered from these instruments are summarized below:

- **Identify the purpose of conducting classroom observations.** Identifying the purpose of a classroom observation exercise guides and informs decisions regarding all aspects of instrument development and use. This includes: instrument content and structure; observer profiles and the training and support they will need; the instrument medium (e.g., paper or electronic); and dissemination and use of information collected. Therefore, identifying the purpose is a critical step in planning for classroom observation use.

- **Engage diverse stakeholders and technical experts.** Collaboration and coordination of multiple stakeholders and experts throughout the entire planning and implementation process is essential to the development of a quality instrument and a successful classroom observation exercise. While the specific purpose and context will determine which individuals to involve, it is critical to include the key intended users of the data. Typical users and stakeholders included are: education officials at national and sub-national levels; reading and language specialists; individuals with expertise and responsibility for providing teacher professional development; research, survey design and education measurement specialists; and program monitoring and evaluation specialists.

- **Develop or adapt a high-quality instrument.** Instrument development and adaptation is an iterative process. To begin, programs either need to select an existing instrument or develop a new one that meets criteria for the intended purpose. Key criteria to consider include: observer capacity; instrument validity and reliability; instrument focus and information needs; item and response formats; instructional categories and behaviors; and feasibility and usability. If an instrument has not been designed with each of these criteria in mind, the instrument may need modifications. If changes are needed, instrument developers must verify that these changes still meet the relevant criteria for validity and reliability; align with the instrument’s intended purpose; and are appropriate for the context, language and observer skill level.

- **Pilot a draft instrument.** Piloting—or trialing a draft version of an observation tool in a setting similar to the setting in which it will be used—is a critical step to verify its validity and reliability. Piloting is also an opportunity to gather feedback from observers on the instrument and to assess the quality and consistency
of observer responses. Piloting should be conducted in a small, targeted sample of schools with observers who have received some training in how to use the tool. Data gathered during the pilot should be analyzed to verify whether the instrument is gathering the desired data. Piloting results should then be used to further modify the instrument and inform observer training.

- **Develop protocols for the classroom observation.** Instruments need to be accompanied by a protocol or guide describing how to appropriately use the instrument and conduct classroom observation visits. The protocol should include clear explanation of instrument items and how to record information. Where relevant (e.g., when observers are instructional coaches), protocols may include guidance on how to use the instrument to inform and provide feedback to teachers.

- **Identify, train and support classroom observers.** Those identified to serve as observers will need training on how to appropriately conduct observations and, in some cases, how to analyze and use the information collected. Training content should align with the purpose of the observation exercise and observers’ skill level and include multiple opportunities for observers to practice using instruments and to be assessed on their ability to administer the instrument correctly, accurately and reliably. If the observer provides feedback to teachers, training also needs to include instruction on how to use information to inform feedback. Staggering the content of training over multiple workshops allows observers time between sessions for practice and skill acquisition.

- **Administer the classroom observation instrument.** Important issues to consider related to instrument administration include the medium of the instrument (i.e., paper or an electronic device such as a cellular phone or tablet) and the frequency and duration of classroom observation visits. Other key considerations include the following: the instrument purpose; context; observer skill level; cost of hardware, software and technical support; data management and storage; and sustainability, if relevant. While classroom observations are being administered, observers need to be monitored to verify they are administering the instrument as intended and address accordingly.

- **Disseminate and use observation results.** Information collected from classroom observation instruments is only useful if it is appropriately shared and used. When identifying when, how and with whom to share findings, the following factors need to be considered: purpose and audience for information sharing; strategies for communicating findings; frequency for disseminating findings; and human and other resources needed to share results.

This Toolkit provides recommendations based on the available research on effective design and use of classroom observation, and current experiences using observation to support EGR improvement. Additional research, reflection and collaboration are needed on the following: (1) improved planning to allow for sufficient time to develop and pilot instruments; (2) improved quality assurance to better align instruments with their purpose and to make them more reliable, valid, user-friendly and feasible to implement; (3) expanded focus of classroom observation to move beyond teachers’ adherence to specific lesson plans to include more emphasis on the quality of teaching and learning; and (4) increased sharing of experiences to generate more best practices and tools.
Introduction

A key strategy of the U.S. Agency for International Development (USAID) Education Policy (2018) for achieving “sustained, measurable improvement in learning outcomes and skills development” is to transform teacher professional development systems to improve classroom instruction. One way to improve teacher professional development is to observe teachers in their classrooms with the aim of helping them improve their practice. The use of classroom observations can serve different purposes, from informing the design of professional development initiatives to providing individualized feedback to teachers to evaluating the effectiveness of a new approach to instruction.

A review of practice to date in USAID-supported early grade reading improvement programs reveals a need for better quality and more purpose-oriented observation practices to help improve teacher performance in the classroom. This Classroom Observation Toolkit for Early Grade Reading Improvement therefore provides guidance to diverse stakeholders using teacher observation around the world on the principals of high-quality classroom observation in primary grade reading instruction, and supplies practitioners with guidelines, examples from practice and observation templates that can help strengthen the use of observations in early grade reading programs.

The intended audience for this Toolkit includes USAID personnel, governments collaborating with USAID to improve early grade reading outcomes, and USAID’s diverse implementing partners.

The Classroom Observation Toolkit is organized as follows:

- **Understanding Classroom Observation** provides a general overview of classroom observation and observation instruments. Informed by a review of the relevant literature, the purpose of this section is to define classroom observation, its importance, and the different purposes for which it is used. It also describes observation instruments and key features of instrument content and structure.

- **EGR Program Experiences and Guidance** shares EGR programs’ experiences developing and using classroom observation for different purposes. Based on responses to a survey on EGR programs’ experiences using classroom observation, as well as a review of existing instruments, the purpose of this section is to summarize EGR programs’ experiences in the following areas: purposes and use of observations; features of classroom observation instruments; instrument development and adaptation; preparation and training of observers; instrument administration; and dissemination and use. The purpose of this review is to understand EGR programs’ current challenges, successes and needs with respect to classroom observation. Following the summary of experiences, the section provides guidance on each topic based on lessons learned from practice and best practices as identified in the literature.

- **Summary and Future Directions** summarizes key takeaways based on review of EGR experiences and instruments. It describes areas where the EGR community can collaborate to improve classroom observation instrument development, implementation and use in early grade reading programs.

This resource also includes five **Annexes to support instrument development**. These include a checklist of key considerations for instrument development and four observation instrument templates tailored for different purposes. The templates illustrate potential formats and content depending on the context and purpose of the observation.

Information and guidance presented in this toolkit is informed by relevant literature on classroom observation and program reports, a review of available instruments and resources, and information and experiences gathered from a Global Reading Network survey about classroom observations in EGR programs. While it is not a comprehensive guide on all aspects of classroom observation, it provides practical information and examples from current program implementation to support a broad audience in the effective use of classroom observation.
Understanding Classroom Observation

The purpose of this section is to provide a general introduction to classroom observation and instrument content and structure. Specifically, this section of the toolkit:

- Defines classroom observation;
- Explains the importance of classroom observation;
- Describes general purposes and uses of classroom observation; and
- Describes key features of instrument content and structure and considerations for different purposes.

The content of this section is informed by a desk review of current literature on effective classroom observation and instrument development. Key terms and concepts identified in this section provide the overarching framework that guided the development of this toolkit including the analysis of EGR program experiences and guidance provided in the later section on EGR Program Experiences and Guidance.

Defining Classroom Observation

Classroom observation is the process of an individual observing a teacher facilitating a lesson either in a classroom or another type of learning environment. During the observation, the observer typically uses a tool or instrument to record what he or she sees in terms of teacher behavior, student behavior, resource use, and/or the classroom environment itself.

While classroom observations are a common practice in most educational settings, the focus and content of the observation varies depending on the purpose of the observation and intended use of the data gathered. Classroom observations are sometimes referred to as teaching practices inventory, teacher and student observation, or classroom or literacy walkthroughs. This Classroom Observation Toolkit uses the term classroom observation to be inclusive of all types of observations conducted by EGR programs that include a focus on teacher behavior and classroom instruction.

Importance of Classroom Observation

Research has demonstrated the critical role of teachers and the important influence they can have on student learning. Teachers’ ability to use specific teaching practices with their students matters. Research also has identified specific characteristics of effective teacher professional development that help facilitate teachers’ ability to learn and be able to apply specific teaching practices and behaviors. Included among these characteristics of effective professional development are continuous feedback, reflection, coaching and ongoing support to teachers. Classroom observations play an important role in each of these professional development efforts as they help inform teacher feedback and tailor further professional learning opportunities.

---

Evidence from research has also demonstrated the important role classroom observations have in improving student learning. A large-scale research study in the United States analyzed data collected from five different classroom observation instruments and found significant correlations between documented evidence of improved teaching practices and improved student learning outcomes. Additionally, results from recent studies comparing the use of specific observation instruments in low- and middle-income countries found similar correlations between observed teaching practices and higher student learning gains.

Purpose and Use of Classroom Observation

The overall goal of any classroom observation is to improve the quality of teaching and learning in the classroom, while the purpose of conducting a specific classroom observation and the use of the information collected can vary. This section highlights some of the common purposes and ways identified in the literature in which classroom observations can support teacher professional development and improvement in student achievement.

Identify Teacher Professional Development Needs

Classroom observations can help identify teacher professional development needs and inform the design of a new intervention or training. Classroom observations are just one aspect of a comprehensive needs assessment process. They serve to clarify existing positive teaching practices and behaviors and identify ongoing professional development needs. Other aspects of a needs assessment collect information on teacher attitudes, beliefs, and pedagogical and content knowledge. In contexts affected by crisis or conflict, observations might be used as part of a rapid assessment that assesses different aspects of the educational context including a review and analysis of the curriculum, availability and priorities of existing training structures, or availability and need for psycho-social support to teachers and students, etc. (see USAID’s Rapid Education and Risk Assessment Toolkit and INEE’s Guidance Note on Conflict Sensitive Education). Typically, observers who conduct observations as part of a needs assessment are individuals who support the initiative. However, the observer may vary depending on the type and scale of the initiative. Information collected during a needs assessment can be used to design a professional development intervention tailored to the specific needs of a group of teachers.

---


Support Individual Teacher Professional Growth and Learning

Evidence in the United States has consistently identified “job-embedded learning” and supportive feedback and reflection as essential for teacher growth and learning. Research has also demonstrated the critical importance of teachers having sustained, on-going support through coaching or mentoring or professional learning communities. Recent research from Brazil found that feedback, when combined with on-going coaching or mentoring, was a “promising strategy for whole-school efforts to raise teacher effectiveness.” Professional development programs that uphold these evidence-based practices utilize classroom observation as an improvement-oriented activity within a larger professional development effort to help both the teacher and the observer reflect on an individual teacher’s strengths and needs, establish goals for improvement, and identify areas where additional support is needed.

Observers for this purpose are usually instructional specialists, trainers, coaches, principals, or peer teachers that work at the school level and have established a sense of trust with teachers. The observation provides information that is shared as feedback with the teacher in a targeted session. The purpose of the feedback session is to prompt self-reflection, engage teachers in dialogue about their practice, empower them and enhance learning, thus stimulating change. During the feedback session, the observer and teacher reflect together and identify improvement goals and the support needed by the teacher to achieve those goals. The observer then provides on-going coaching or mentoring support to help teachers achieve their improvement goals.

Monitor Program Implementation

When a teacher professional development initiative introduces teachers to a new approach or materials, classroom observations can serve as a monitoring activity to verify whether teachers are using newly introduced strategies or materials as expected. Information learned from monitoring observations helps a program understand whether teachers are implementing with fidelity, and whether teacher behavior is changing over time.

Observers who conduct program monitoring observations are typically individuals who are responsible for supervision and monitoring of the initiative rather than individuals who provide direct technical assistance to the program. Programs use classroom observation as one source of data collection in conjunction with other monitoring data sources to understand the extent to which an intervention is making progress as measured by its indicators and where adjustments are needed. Government supervisors, donors or program advisors may also conduct classroom observations to help monitor implementation and identify modifications needed in the training or program design. For example, program advisors or managers may use observation data and compare with pre- and post- training test data to understand to what degree linkages exist between teacher knowledge gained through training and changes in teacher practice. Midway through and at the end of an initiative, observation data, along with other monitoring data, are compared with baseline data and analyzed and interpreted and compiled into a mid or final evaluation report, outcome, or performance evaluation.

Evaluate the Impact of Professional Development

Classroom observations also serve to provide data on the impact of a professional development intervention on improved teacher performance and student achievement. The focus of the observation will vary depending on how the intervention defines impact. For example, some reading improvement programs might define impact as an increase in the

---


9 See, for example, Darling-Hammond, Hyler, & Gardner, Effective Teacher Professional Development.

10 See, for example, Bruce R. Joyce, and Beverly Showers. Student achievement through staff development (3rd ed.) (Alexandria, VA: Association for Supervision & Curriculum Development (ASCD), 2002); Darling-Hammond, Hyler, & Gardner, Effective Teacher Professional Development.


13 For more information on coaching in EGR programs, see Marion Fesmire and Amy Pallangyo. “Teacher professional development and coaching in early grade reading programs.” [Webinar]. Presented by Global Reading Network in Washington, DC, November 15, 2018. For information on peer coaching in crisis contexts, see INEE’s Teachers in Crisis Contexts (TICC) Peer Coaching Pack.

percentage of teachers implementing specific instructional practices introduced in the professional development program. They might also be interested in understanding the impact on the amount of instructional time teachers use teaching specific reading skills and the frequency of student exposure to those methods. Others might be interested in measuring the impact on the quality of teaching practice.

Evaluations that measure impact of professional development follow specific protocols and thus require a more rigorous process in sampling teachers, instrument development and administration, and training of observers. For example, measuring the impact of a program requires the use of a comparison group to conclude that observed changes are a result of the intervention. The observers for impact evaluations are typically individuals trained in data collection. Some research designs aim to identify if a correlation exists between certain teaching behaviors and practices that lead to improved student achievement and thus require specific data analysis techniques to compare data and find these correlations. Results of impact evaluations or research are shared in reports with donors and policy-makers as evidence to inform scale-up of a program, policy reform, and/or decisions about funding.

Measure and Improve the Quality of Teaching Instruction at a System Level

Classroom observations can also be used to measure and improve teacher quality at a system level. Observations for these purposes are high-stakes and seek to collect objective, quantitative, comparable data to inform system-level decision-making about policies and initiatives implemented at scale. Currently, observations for this purpose are more common in developed countries than in lower- and middle-income contexts (LMIC). This is in part due to the multiple challenges of standardizing quality across diverse contexts. However, efforts to understand how observations can collect comparable data of teacher performance in LMIC contexts are increasing.

Observers are often data collectors hired from local data collection firms or ministry of education personnel who participate in a specialized training. However, given the large number of observers required and the high costs associated with observations conducted at scale, some assessments obtain videos of teaching practices and the observers watch the videos rather than visiting individual classrooms.

Generate New Knowledge About Teaching-Learning Processes

Classroom observations may also focus on learning behaviors and processes to make claims about teaching and learning in a specific context. For example, an academic scholar might be interested in studying patterns in student behavior in both rural and urban contexts in Uganda. Or a government may want to conduct research to explore teacher attitudes and behaviors that correlate with/lead to decreased student drop-out rates. An early grade reading program may want to conduct classroom observation to understand the read-aloud practices of teachers or the use of locally produced materials in rural classrooms. Classroom observations for these purposes are not meant to assess an individual teacher or program, but rather to generate new knowledge about teaching and learning in a specific context.

Evaluate Individual Teacher Performance for Accountability

Many educational systems conduct classroom observations to evaluate individual teacher performance on a specific set of standards established by government. Observers might include inspectors, administrators, supervisors, principals or other school leaders responsible for observing teachers on a regular basis and holding them accountable for meeting specific standards. Classroom observations are usually one component of a larger evaluation system to evaluate teacher performance. Educational systems and schools use results from this type of

---

15 Sarah Pouezevara et al. Measures of quality through classroom observation for the Sustainable Development Goals.
16 Ibid.
17 Ibid.
18 Ibid.
20 See, for example, Megan Stuhlman, Bridget Hamre, Jason Downer, and Robert Pianta. A Practitioner’s Guide to Conducting Classroom Observations: What the Research Tells Us About Choosing and Using Observational Systems to Assess and Improve Teacher Effectiveness (Charlottesville, VA: Center for Advanced Study of Teaching and Learning, University of Virginia, 2014).
observation data set to inform decision-making about an individual teacher’s employment, compensation, promotion, or other types of incentives.

For more information on purpose and use of classroom observations in the context of EGR programs, see *Purpose of Classroom Observations in EGR Programs* in the section *EGR Program Experiences and Guidance*. This section describes how reading programs have used classroom observation for the purposes described above, as well as key considerations related to identifying how classroom observation can be used.

**Instrument Features and Content**

During a classroom observation, observers must use a “well-developed” instrument to document what they see taking place. Instruments help focus the observation and gather useful information to help improve teacher practices and ultimately student learning outcomes. An extensive amount of literature exists on best practices to develop instruments. This section highlights key features of observation instruments commonly discussed in the literature as important.

**Validity and Reliability**

A well-developed instrument has evidence of validity and reliability. Validity is the extent to which an instrument measures what it aims to measure. A classroom observation instrument that is valid captures instructional practices and behaviors that lead to improved student achievement. Reliability refers to whether the instrument consistently measures what it intends to measure. In other words, the information collected on an observation instrument during a single observation is consistent across observers.

Instrument validity and reliability are particularly important when an observation instrument aims to measure or assess the quality of teaching instruction or collect comparable data. To determine an instrument’s validity requires rigorous review and statistical assessments conducted by subject-matter experts and researchers. Verification of an instrument’s validity are typically provided in documentation that accompanies the instrument. To determine an instrument’s reliability also requires rigorous testing of an observer’s ability to score consistently on a given item over time as well as their ability to score consistently with others. Such assessments are conducted by researchers and evaluators as well. However, even if an instrument is verified as reliable and valid in one context, it must still be verified within the setting it will be used. For example, to ensure the reliability of an instrument, observers must be assessed for what is referred to as *inter-rater reliability* (IRR). Inter-rater reliability refers to the degree or extent to which a group of observers (or “raters”) scores a given item the same. IRR assessment typically occurs during observer training on the observation instrument. Training that results in high IRR scores is more likely to lead to high-quality and reliable data that is accurate and consistent across raters.

**Item and Response Formats**

Items on an observation instrument are questions or statements about the specific actions, behaviors or processes being observed. Response formats are the way in which information for each question or item is recorded or documented. The item and response formats on an instrument will vary depending on the purpose of the observation, what it is measuring, and the level of inference or judgment the observer must make when recording information.

If the purpose of the observation is to understand teachers’ adherence to specific instructional strategies or use of instructional time and the observer needs to identify specific instructional practices that occurred, the instrument would typically include low-inference items, or behaviors that are distinct, observable and objective, requiring no evaluative judgment by the observer. An example of a low-inference statement is “The teacher used at least two...”

---


23 Ibid.

different teaching and learning materials to support the lesson.” Low-inference items help maximize an instrument’s potential for greater reliability.

If the purpose of the observation is to assess or make a judgment on how often or how well teachers used instructional strategies, the instrument would use high-inference items, which are subjective. An example of a high-inference item is “Teacher used teaching and learning materials effectively.” This statement requires the observer to make a judgment about whether the teacher used the materials effectively, which requires the observer to understand what effective use of material looks like in the classroom.

See Table 1 for more examples of low-inference and high-inference items.

**Common item response formats** on observation instruments include the following:

- **Binary response:** Item format that provides the observer a choice of two responses, such as yes/no. The binary response format provides information as to whether a specific teacher behavior is observed or not.

- **Time-interval sampling:** Item format that allows the observer to record specific behaviors or processes and the frequency of those behaviors in a specified time interval. This format can help the observer capture teachers’ use of instructional time or amount of time students are engaged.

- **Rating scales:** Item format that uses numerical numbers for scoring responses (1=never, 2=sometimes, or 3=always) or text responses (never, rarely, sometimes, frequently, and always). The rating scale format allows the observer to rate the frequency or quality of specific items observed in the classroom. This requires the observer to make a judgment using specified criteria on a rating scale.

The most common response format in instruments with low-inference items involves binary responses where the observer selects “yes” if a specific behavior was observed during a lesson or “no” if it was not. See Annexes B, C, D and F for examples of instruments with low-inference items and binary response format. Low-inference items can also be used with a time-interval sampling response format to capture teachers’ use of instructional time on specific behaviors. See Annex E for an illustrative example of an instrument with low-inference items and a time-interval sampling response format.

The most appropriate response format for high-inference items is a rating scale where the observer scores teacher performance using numbers (i.e., 1-4) or a descriptive rating scale (i.e., never, sometimes, consistently). Rating scales can be difficult for some observers to score reliably and thus, require very specific criteria for scoring and an adequate amount of training to achieve reliability across observers.

**Focus of Instrument**

The overall focus of an observation instrument varies depending on the purpose of the observation and what it aims to measure. This section describes the focus of classroom observation instruments by the different observation purposes commonly discussed in the literature.

- **Identify teacher professional development needs:** The focus of observation instruments that identify teacher professional development needs prior to developing an intervention can vary depending on the context and specific interests of the needs assessment. If the interest is to inform the design of an EGR intervention, the instrument might collect information on teachers’ use of evidence-based instructional strategies and literacy practices. The instrument might also collect information on the availability of materials in the classroom, as well as teacher and student use of materials in the classroom. In contexts where an EGR intervention may aim to improve social and emotional learning, an instrument might include additional items relevant to teaching strategies that support students’ emotional health and well-being. (This may be particularly important in contexts affected by conflict and crisis.) Classroom observation instruments used for needs assessment purposes are typically used in combination with other data collection tools that aim to understand teachers’ knowledge and beliefs about using teaching and learning materials or their perceptions of their use of materials in the classroom.

- **Improve teacher professional growth and learning:** Classroom observation instruments used to
improve teacher professional growth and learning help identify an individual teacher’s strengths and needs and prompt reflection on improvement goals. Observation instruments for this purpose can be either unstructured or structured. An unstructured instrument does not have specific criteria; it is more open-ended and allows for observers to take notes describing what they see. Structured instruments target specific instructional practices or competencies to be observed such as student-teacher interactions, use of teaching and learning materials, questioning patterns, etc. Structured instruments help focus observers on specific items for which they need to record information. Recording evidence or examples on either type of instrument is an important aspect of any observation used to inform feedback and improve individual teacher growth. It provides the observer and teacher with concrete examples to refer to during the post-observation discussion when identifying strengths and areas of improvement. Instruments may also include documentation of the results of feedback and include improvement goals discussed between the teacher and observer and what type of support the teacher needs.

- **Monitor program implementation**: Observation instruments for monitoring implementation will vary based on a school or program’s monitoring needs. For example, if a classroom inventory is needed for procurement purposes, the instrument may capture information regarding books and supplies in the classroom. Classroom observation instruments that aim to monitor fidelity of implementation (FOI) might include content that focuses on monitoring adherence, duration/exposure, use of instructional time, quality of delivery, student responsiveness, or other program specifications. However, in the context of EGR programs the most common and simplest of FOI instruments focus on adherence to guided lesson plans, use of teacher and learning materials and duration or frequency of student exposure to specific instructional strategies. For this, the scope and content of the instrument directly align with what the program expects the teacher to be doing for a specific lesson or any given period.

- **Evaluate program impact**: Instruments for evaluating program impact focus on teacher behavior targeted in the intervention and track changes in behavior over time. The content of classroom observation instruments used for evaluating impact will differ depending on how the program defines impact. Instruments can be developed to evaluate impact of the program on teacher adherence to teacher strategies, duration/exposure, use of instructional time, time on task, quality of delivery, student responsiveness, or other interests of the program. If the purpose is to measure the quality of teacher performance on behaviors promoted in the program, the content of the instrument will focus on program-specific behaviors. If the purpose is to measure the impact of professional development on teacher quality, the content of the instrument might not be program-specific and will include general measures of quality evident in standardized observation instruments. Observations for measuring program impact are typically conducted in a sample of intervention schools as well as a sample of control group schools not supported by the program to isolate program impact.

- **Measure and improve quality of instruction at systems level**: Instruments used to measure and improve quality of instruction at scale provide reliable quantitative measures of classroom instruction proven to lead to student achievement. Instruments focus on general teacher and student behavior and practices that can be observed across multiple schools, programs or countries and compared globally. The most common instrument validated in several LMIC contexts is the Stallings classroom snapshot instrument. Stallings uses

---

26 Ibid.
27 Ibid.
28 Bruns, Barbara et al. *Measures of Effective Teaching*. 

---

10 Classroom Observation Toolkit
time-interval sampling methodology to capture student-teacher interactions, information on teachers’ time on instruction, use of different activities, and ability to engage students.\(^{29}\) While programs use it to evaluate the impact of their program on the quality of teacher instruction, it is not an instrument that reflects a specific reading curriculum. Another example is an instrument recently piloted and soon to be launched by the World Bank referred to as the TEACH\(^{30}\) instrument, an open-source instrument specifically developed for use in low- and middle-income countries. Both the Stallings and TEACH instruments have consistently produced reliable data on the quality of teacher instruction as measured by student-teacher interactions.

- **Evaluate individual teacher performance for accountability:** Instruments that evaluate individual teacher performance for accountability purposes measure teacher proficiency levels on specific performance standards or competencies. The observation instruments align with the specific performance standards or competencies established at the national, district, or state level, which provide teachers with a shared vision of quality teaching they strive to achieve.

**Instructional Categories and Observable Behaviors**

High-quality classroom observation instruments are organized into instructional categories or dimensions that have been proven to lead to student achievement. Instructional categories proven as predictors to student learning across different contexts and included as measures in well-known standardized observation instruments\(^{31}\) include the following categories or some variation of them:

- Lesson structure, content and facilitation (e.g., explicit reading instruction)
- Classroom management
- Physical/classroom environment
- Classroom culture/supportive learning environment
- Student participation/engagement
- Checking for understanding/assessment
- Feedback

Each category of instruction comprises several sub-categories of behaviors that may define the overarching category or serve as indicators of success for that measure. For example, lesson strategies and content delivery may include behaviors that indicate teacher preparedness, use of lesson plans, use of content-specific instructional strategies, use of teaching and learning materials, and asking questions. The extent to which teachers show evidence of performing each of these behaviors as measured by an instrument would indicate to some degree their proficiency in the category of lesson structure, content and facilitation and thus predict their likelihood to impact student learning.

Within the different categories of instruction, instruments provide a list of specific observable behaviors\(^{32}\) or processes that guide what to look for during an observation. Observable practices or processes are distinct items that an observer may see during a lesson relevant to the category it measures. For example, “a teacher prepared a lesson in advance” is not an observable behavior: This is not something the observer can see within the context of the observation. However, the observer can see if “a teacher has a lesson plan” or “states the lesson objective to the class.”

---


\(^{30}\) Molina, Ezquiel et al., Measuring Teaching Practices at Scale.


\(^{32}\) Jilliam Joe et al., Foundations of Observation.
In an instrument, observable practices are written as either low-inference or high-inference questions or statements, as discussed previously. When observable behaviors are written as high-inference, they require specific criteria for scoring that are clearly described in a rubric.

See Table 1 for a list of common categories and sub-categories of classroom observation instruction proven to link to student achievement across many contexts with examples of low-inference and high-inference observable practice statements.

**Feasibility and Usability**

Instrument feasibility and usability are important to consider when developing or selecting instruments. Feasibility refers to the potential of an instrument to be used appropriately and effectively given the purpose, scope of the instrument, context, observer capacity, and time and period in which the observation will be conducted. For example, sometimes an instrument that might be appropriate in one context might be too difficult, too costly or too lengthy when used in a different context with a different budget, different observers or different amount of time for administering the instrument. Usability refers to the degree of “user-friendliness” of the instrument itself as well as the processes for data management and analysis. A user-friendly instrument that is paper-based requires a format, structure and font that are easy to understand and use and an easy-to-code format for data entry. For example, an instrument that has small font and large amounts of text may be challenging for the observer to read; instruments with inconsistent formats throughout the document which require observers to shift their approach each time might also be difficult. Binary checklists provide an easier-to-code format and data management process than other types of instruments. If instruments are not both feasible and user-friendly, they will not be effective or reliable.

For more information on instrument features and content in the context of EGR programs, see Observation Instrument Features and Content, in section EGR Program Experiences and Guidance for a summary of experiences to date and guidance on improving these aspects of classroom observation instruments.

---


Table 1. Instructional categories and examples of observable behaviors

<table>
<thead>
<tr>
<th>Instructional Categories</th>
<th>Instructional Sub-categories</th>
<th>Low-Inference, Observable Behaviors (Example)</th>
<th>High Inference, Observable Behaviors (Example)</th>
</tr>
</thead>
</table>
| Lesson structure, content and facilitation | ● Teacher preparedness (lesson planning and objectives)  
● Language of instruction  
● Use of content-specific instructional strategies (i.e., foundational skills for reading\(^{36}\))  
● Use of general instructional strategies (i.e., structured pedagogy principles, social and emotional learning practices\(^{37}\))  
● Use of teaching and learning materials  
● Critical thinking | Teacher used teaching and learning materials in the lesson | Teacher used teaching and learning materials effectively |
| Classroom management | ● Time management (pacing time on task, transitions, etc.)  
● Classroom organization (seating arrangements)  
● Groupwork, pair work, individual work  
● Behavior management (displays rules, use of routines) | Teacher has behavioral expectations displayed on the wall for students to see | Teacher behavioral expectations are clear for students |
| Physical/classroom environment | ● Availability of teaching and learning materials  
● Availability of texts  
● Print-rich environment (displays of text)  
● Safe and inclusive learning environment\(^{39}\)  
● Desk and chairs/sitting space | Teacher has at least 3 print-resources posted on the wall | Teacher has adequate print resources displayed in the room |
| Classroom culture/supportive learning environment | ● Positive, clear communication and attitude  
● Positive discipline and use of praise  
● Equitable and inclusive practice\(^{40}\) | Teacher provides equal opportunities to girls and boys to ask questions | Teacher uses inclusive strategies to engage all students |
| Student participation/engagement | ● Student independent practice  
● Student-to-student interaction  
● Level of participation | Students practice reading or writing independently (not copying) in at least one activity. | Teacher uses effective strategies to engage students |
| Checking for understanding/assessment | ● Monitoring individual and group work  
● Asking questions to whole class and individuals | Teacher walks around classroom and checks all students’ work during independent, pair or group work. | Teacher monitors students’ work adequately |
| Feedback | ● Affirmation of correct responses  
● Clarification/correction of misunderstandings | When students perform incorrectly, teacher corrects responses AND provides an opportunity to try again. | Teacher provides constructive feedback to students |


\(^{37}\) Young-Suk G. Kim, & Marcia Davidson. Promoting successful literacy acquisition through structured pedagogy: Global Reading Network Critical Topics Series (Chevy Chase, MD: Prepared by University Research Co., LLC. (URC) under the Reading within Reach (REACH) initiative for USAID’s Building Evidence and Supporting Innovation to Improve Primary Grade Assistance for the Office of Education (E3/ED)).

\(^{38}\) For resources on SEL learning, see Interagency Network for Education in Emergencies (INEE), Psychosocial Support and Social and Emotional Learning in Emergency Settings (New York, INEE, 2016); International Rescue Committee (IRC), Safe Healing and Learning spaces Toolkit.


\(^{40}\) For more information on promoting literacy with students with disabilities, refer to USAID’s Universal Design for Learning to Help All Children Read (2019). For more information on enhancing girls’ education and gender equality in the classroom, see INEE’s Guidance Note on Gender (New York, NY, INEE, 2019). See also relevant sections of Department for International Development (DFID)’s Girls’ Education and Gender Equality: Education Rigorous Literature Review (2014).
EGR Program Experiences and Guidance

This section describes EGR program experiences and shares practical guidance on the following aspects of instrument development and use:

- purposes of classroom observation commonly used within the context of EGR programs;
- instrument features and content;
- instrument development and adaptation;
- preparation of observers;
- instrument administration;
- dissemination and use of results; and
- planning for classroom observation.

The content of this section is informed by best practices and lessons learned gathered from a combination of a review of the literature, available reports and instruments, and feedback received from the GRN survey on classroom observation use in early grade reading programs. A total of 24 survey responses and 24 classroom observation instruments were reviewed from 13 organizations, representing 24 countries and 29 programs.

Each section includes: 1) a summary of EGR experiences to date with respect to the particular topic, based on the GRN survey and a review of instruments; 2) examples from practice that demonstrate promising strategies; and 3) key considerations and guidance relevant to each topic.

Purpose of Classroom Observations in EGR Programs

**KEY TAKEAWAYS on Purpose of Classroom Observations in EGR Programs**

- EGR reading improvement programs typically use classroom observations for multiple purposes in a single intervention.
- The three most common observation purposes in EGR programs currently include: 1) supporting teacher professional growth and learning through feedback and coaching, 2) monitoring fidelity of implementation, and 3) evaluating the impact of professional development on teachers’ use of instructional time.
- Other less common purposes reported, but equally important to improving EGR teaching and learning, include identifying teacher professional development needs prior to an intervention and generating new knowledge about teaching and learning processes through research.
- Identifying the purpose of classroom observations guides and informs decisions regarding all aspects of instrument development and use. Thus, it is a critical first step when designing and EGR intervention.
Experiences from EGR Programs

This section highlights key findings from the GRN survey on how EGR improvement programs are using classroom observations. Responses to the EGR program survey indicate that many reading improvement programs use classroom observation for multiple purposes within a single intervention. The three most common purposes for which EGR interventions have used classroom observation are to: support teacher professional growth and learning, monitor fidelity of implementation, and evaluate the impact of professional development on teacher changes in instructional practices. However, some EGR programs use observations for other purposes such as identifying teacher professional development needs or exploring teaching and learning processes through research. According to the survey results, classroom observation has been less commonly used to measure the quality of teaching at a system level and to evaluate individual teacher performance for accountability purposes. This section describes EGR programs’ experiences using classroom observation for the most common purposes reported in the GRN survey.

Support Individual Teacher Professional Growth and Learning

Many EGR initiatives use classroom observation to support individual teacher growth and learning through regular feedback and coaching. Observations are part of an ongoing cyclical process that involves a pre-observation meeting, observation, and post-observation discussion. In a pre-observation visit, coaches and teachers meet before a lesson to discuss the upcoming lesson and what the teacher would like the coach to focus on. Depending on the needs and goals of the teacher, coaches may help teachers prepare for the lesson or model specific teacher behavior that will be part of the lesson. The coach then conducts the observation using an instrument that helps document evidence of good practices rather than making judgment about practice. Coaches complete the observation form during the observation and then give feedback after class. Several survey responses emphasized that feedback is “not punitive in any way.” The purpose of the feedback is “to guide and accompany by reflecting on teacher observation, not supervising or auditing to see if the teacher is doing things right.” See Example from practice 1 for more details on how two initiatives use an observation process to support teacher reflection and growth.

Monitor Program Implementation

In addition to coaching, several EGR interventions use classroom observation to monitor implementation and to track and measure progress against program indicators. Many evidence-based reading improvement initiatives monitor fidelity of implementation. This is becoming increasingly common in EGR initiatives due to evidence that implementation fidelity leads to greater program impact. Classroom observations used for monitoring FOI help track teachers’ adherence to the instructional strategies as planned and, in some cases, the frequency, duration,

---

41 See, for example, Bruns et al. Through the Looking Glass: Can classroom observation and coaching improve teacher performance in Brazil?
and quality of implementation. By collecting monitoring data periodically, programs track changes in teacher behaviors over time and identify areas where the program or professional development needs improvement. See Example from practice 2 from an initiative in Madagascar that uses observations to monitor FOI. Additionally, M&E teams use FOI data for reporting on indicators. Results are shared with staff and government officials and used to inform further decision-making about the program.

**EXAMPLE FROM PRACTICE 2**

Classroom Observation for Monitoring

The Madagascar Mahay Mamaky Teny initiative, implemented in 2018 by FHI 360, used classroom observations to monitor the fidelity of implementation to understand whether teachers were following the steps of the lesson plan and applying strategies they learned in the training. Observers were enumerators from the Ministry of Education evaluation team who were involved in assessment and ongoing formative evaluation of the project (GRN survey response). Observations were administered twice for each teacher during a four-month pilot.

**Evaluate the Impact of Professional Development**

Classroom observations used by EGR improvement initiatives also are commonly used to help measure the impact of professional development on teacher performance. Most evaluations used observation to identify changes in teachers’ behavior over time and/or changes in their use of instructional time. In a few programs, evaluations used observations to measure the changes in the quality of teacher practices. Classroom observations were conducted in the same schools at the same time as learning assessments were conducted with students using the Early Grade Reading Assessment (EGRA). In some evaluations, the results of the classroom observation were linked to the EGRA results for children in the classroom observed, a useful approach to understand linkages between specific teacher practices and improvement in student reading skills. See Example from practice 3 to learn how one EGR initiative used classroom observation to evaluate the impact of professional development, as well as to assess the relationship between instructional practices and learning gains.

**EXAMPLE FROM PRACTICE 3**

Classroom Observation for Evaluating the Impact of Professional Development

The Nigeria Reading and Access Research Activity (RARA), implemented by RTI International from 2014-2015, used classroom observation to measure the impact of teacher professional development in improving teachers’ practices as well as how those practices in turn affected student reading gains. The evaluation was informed by the use of both a time-interval sampling instrument and an instrument with a series of low-inference items with binary response formats (e.g., yes/no). The information from the two different instruments provided a complementary and holistic understanding of what was happening in the classroom in terms of teachers’ use of effective instructional practices, as well as how much time they spent teaching key reading skills (RTI International, 2016). The observation data was subsequently linked to students’ scores on a reading assessment to quantify the relationship between changes in teacher instruction and student outcomes. The study found a positive association between the two: Students’ literacy outcomes improved as their teachers’ instruction did. In other words, as teachers adopted more effective instructional literacy practices, children’s reading skills improved Pflepsen, A., Harden, K. & King, S., in press).

---

42 Cao et al. R&E Search for Evidence.


44 Linking classroom observation and student assessment data requires careful advance planning. The section Planning for Classroom Observation: Additional Considerations provides some guidance on this process, but EGR programs are encouraged to consult with the relevant experts (sample specialists, M&E experts) during the design phase of the classroom observation exercise to set it up well.
Other Purposes of Classroom Observation in EGR Programs

EGR programs also use classroom observations for other purposes, though such uses were less common. Of the surveys reviewed, some programs used classroom observation to identify teacher needs to inform the intervention design\(^{45}\) to examine the readiness of an intervention to be evaluated\(^{46}\) (e.g., evaluability assessment), and to explore teaching and learning processes in the classroom through research.\(^{47}\) (For example, Example from practice 4 describes how classroom observations have been used to support research in initiatives in South Africa and Nigeria.) Although the GRN survey responses indicated that classroom observation was not commonly used in participating EGR initiatives for these less common purposes, EGR programs are encouraged to use classroom observation for these other purposes (e.g., as part of a situational or contextual analysis prior to designing an intervention to understand what the current practices and needs are in the classroom and what type of professional development is needed).\(^{48}\)

---

**Example from Practice 4**

Classroom Observation to Support Research

In South Africa, the Department of Basic Education (DBE) conducted a large-scale Early Grade Reading Study (EGRS) (implemented from 2015-2020) to quantitatively assess the impact of an early grade reading intervention on student learning. To complement this study, the DBE also conducted a separate study called the Classroom Observation Study in a sample of successful intervention schools to gain an in-depth understanding of changes in teaching practices, as well as to identify how the professional development intervention contributed to those changes. Classroom observations represented one of three data collection methods and helped to gather qualitative data on the teaching and learning environment. Information collected provided rich data to support findings from the EGRS impact evaluation (Department of Basic Education, 2017).

In another research study, in Northern Nigeria, classroom observation data has been used to better understand the relationship between coach characteristics, the quantity of coaching, and changes in teachers’ instructional practices. Grade 2 Hausa teachers’ reading instruction was observed prior to and immediately following a pilot reading intervention, conducted as part of during the Reading and Access Research Activity (RARA), an initiative funded by USAID and implemented by RTI International from 2014 to 2015. The observation data, which showed improvement in teachers’ instruction over time, was analyzed vis-à-vis their coaches’ background and experience and the quantity of coaching they received (e.g., number of visits). The findings from the study provide insights that can be used to inform coach recruitment and training, as well as the frequency of coach visits. The research also pointed to a need to better assess the quality of instruction through classroom observation (Harden, King, and Pflepsen, 2019).

---

**Guidance on Instrument Purpose and Use**

Identifying the purpose(s) for using classroom observation is critical during the early stages of deciding if and how to use classroom observation. Key recommendations regarding the identification of the instrument purpose and how information will be used include:

- **Identify the purpose of classroom observation.** Identifying the purpose will help guide other decisions regarding the appropriate profiles of individuals who conduct the classroom observations, the appropriate

---


\(^{47}\) See, for example, Department of Basic Education, Republic of South Africa. Summary report: Results of Year 2 impact evaluation. The Early Grade Reading Study (EGRS) (Pretoria, SA: Department of Basic Education, RSA, 2017).

\(^{48}\) USAID’s forthcoming Literacy Landscape Assessment tools and guidebook, developed by the REACH initiative in collaboration with the Global Reading Network, describe how observation can be used during the program design phase. The LLA will be available on the USAID Education Links website (www.edu-links.org). Also, the INEE Guidance Notes on Teaching and Learning provides guidance for conducting comprehensive assessments concerning INEE Minimum Standards for teacher training, professional development and support in contexts affected by emergencies, chronic crises and early recovery.
medium of the instrument, the processes that will be used for information flow, the resources needed to develop and administer classroom observation instruments, the content and structure of the instrument one might need, and the dissemination and use of information collected. See Table 2 for examples of observation purposes relevant to EGR initiatives.

- **Decide how the program will use the information.** Will data be collected to identify teacher practices and professional development needs prior to an intervention? Will data be collected to inform feedback with teachers and tailor individual support to them or will it be shared internally for monitoring purposes to help track implementation and inform decision-making about the program? Will information be shared publicly with Ministry officials and local stakeholders? Making decisions early on about how the program will use the information will help determine what resources and costs may be needed for dissemination and use. See Dissemination and use of results for additional guidance.

### Table 2. Classroom observation purposes common in EGR initiative

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Description</th>
<th>When conducted</th>
<th>Observer profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify teacher professional development needs</td>
<td>Part of needs assessment or situational analysis to understand teacher beliefs, attitudes, knowledge, and behaviors. Used to inform content of professional development program.</td>
<td>Prior to a professional development intervention</td>
<td>Program personnel or trained data collectors</td>
</tr>
<tr>
<td>Support individual teacher professional growth and learning</td>
<td>School-based activity that helps understand an individual teacher’s strengths and needs to inform supportive feedback to teacher. The purpose of feedback is to prompt self-reflection, engage teachers in dialogue about their practice, empower teachers, and stimulate change.</td>
<td>Ongoing basis</td>
<td>Instructional specialists, coaches, head teachers, peer teachers who support other teachers</td>
</tr>
<tr>
<td>Monitor implementation</td>
<td>Monitoring activity to confirm the program is going as planned and leading to progress on project indicators. Reported in internal monitoring and performance evaluation reports to track progress towards program indicators.</td>
<td>Ongoing basis to allow for adjustments; reported in internal monitoring and performance evaluation reports</td>
<td>Program and M&amp;E personnel and/or MOE personnel responsible for supervision and monitoring</td>
</tr>
<tr>
<td>Evaluate the impact of professional development intervention</td>
<td>Impact evaluation to understand the impact of a professional development intervention on longer-term outcomes (i.e., student achievement, teacher performance). Requires use of a control group and higher levels of reliability in observation instruments. A standardized, validated instrument is used, usually in combination with student achievement scores.</td>
<td>Before intervention begins and after it ends</td>
<td>Researchers and trained data collectors or data enumerators</td>
</tr>
<tr>
<td>Generate new knowledge about teaching and learning processes</td>
<td>A formal or informal research activity to explore specific teaching and learning behaviors and processes. Information used to make claims about teaching and learning in a specific context.</td>
<td>Determined by research</td>
<td>Researchers and trained data collectors or enumerators</td>
</tr>
</tbody>
</table>
Observation Instrument Features and Content

**KEY TAKEAWAYS on Observation Instrument Features and Content**

- The content and features of observation instruments differ across purposes and programs.
- Instrument validity and reliability are a significant challenge in EGR programs and are critical features to consider when selecting or designing an instrument.
- The focus of a high-quality instrument must align with what it aims to measure.
- The item and response format of an instrument must be appropriate to the purpose of the observation, what information is needed and how it will be used, the expertise and skills of those designing the instrument, and the skills and capacity of the observer.
- Instructional categories and behaviors must be evidence-based and align with the content of the professional development intervention and the purpose of the observation.
- A well-designed instrument is user-friendly for the observer and feasible to implement.

**Experiences from EGR Programs**

The types of instruments EGR programs use vary significantly across purposes and programs. Additionally, some programs use individual instruments for each type of purpose while others use one instrument to serve multiple purposes. This section highlights instrument features and shares relevant findings from a review of available EGR classroom observation instruments. These experiences inform the recommendations provided in the section **EGR Program Experiences and Guidance**.

**Validity and Reliability**

Survey results and a review of instruments revealed a need for significant improvement in instrument validity and reliability. Only a limited number of the instruments reviewed had documented evidence for validity and reliability, evidence that is required for instruments used for monitoring and evaluation purposes. Of the available documented evidence, few organizations reported having high IRR scores for their observer cohorts. Most survey respondents reported challenges to achieving high IRR on instrument scoring. Many respondents stated this was due to the disconnect between the instrument and capacity of the observer. However, low IRR rate could also be due to several other factors including poorly constructed items that decreased an instrument’s reliability; use of highly subjective observation items without specific and clear criteria; inadequate training of observers; and inappropriate or poor quality testing for IRR. Reporting on instrument validity was also less evident in the reports reviewed. See **Example from practice 5** to learn about one organization’s experience adapting and testing a classroom observation instrument for validity and reliability for use across different country and programmatic contexts.

**EXAMPLE FROM PRACTICE 5**

**Achieving Validity and Reliability**

As part of the Education in Emergencies: Evidence for Action initiative (2015-2020), the International Rescue Committee (IRC) has been adapting and testing an instrument for Teacher Classroom Observations (TCO) used in program evaluations across several IRC programs to show progress on indicators. The Teacher Classroom Observation instrument measures the quality of teaching practices (specifically, general pedagogy and classroom management), as well as the quality of implementation. Currently, the instrument is used in six different IRC countries and has been revised several times. IRC has assessed the instrument’s inter-rater reliability across countries and found it to vary across contexts. The instrument was used to support evaluation of the Pakistan Reading Project (2013-2020), where it was found to have good reliability and evidence of validity (GRN survey response, 2018).
Item and Response Format

Most observation instruments in EGR programs included a combination of both high- and low-inference items. Only a few instruments had only low-inference questions, while one evaluation instrument used all high-inference questions.

The most common response format in the instruments reviewed was a binary (yes/no) list of observable behaviors, likely a reflection of the low observer skill level and relatively short amount of time available for training that was also reported. However, variation in question formats existed depending on the purpose. For example, most coaching instruments contained binary checklists with space for open-ended notetaking. See Figure 1 for an example from the Northern Education Initiative Plus (NEI Plus) program in Nigeria.

Figure 1. Binary checklist with space for coaches to record comments and evidence

<table>
<thead>
<tr>
<th>LESSON CONTENT: What is the teacher teaching?</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each item describes a teacher behavior or classroom activity. Check YES if the behavior or activity is observed at least once. Check NO if it is not observed, or leave the YES box blank.</td>
<td>Write constructive comments on what you observe.</td>
</tr>
<tr>
<td>1. Does the teacher have the required materials necessary to teach today’s lesson?</td>
<td>☐ Yes ☐ No</td>
</tr>
<tr>
<td>- The teacher has Mu Karanta! Teachers Guide, Pupils’ Book and teaching aides (e.g. chart, flash cards, drawings).</td>
<td></td>
</tr>
<tr>
<td>2. Do all pupils have their Mu Karanta! Pupils book?</td>
<td>☐ Yes ☐ No</td>
</tr>
</tbody>
</table>

Most instruments used for monitoring and/or coaching included binary checklists using a yes/no format. See Figure 2 for an example of a response format from the Philippines Basa Pilipinas early grade reading initiative.

Figure 2. Binary checklist for monitoring adherence

<table>
<thead>
<tr>
<th>Lesson Preparation</th>
<th>Yes, observed</th>
<th>No, not observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The teacher has a lesson plan.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

However, in Malawi, the MERIT program used an observation instrument with a checklist of three options: Yes, Partially, and No. See Figure 3 for an example from the MERIT program.

Figure 3. Checklist with three different options

<table>
<thead>
<tr>
<th>Part 2: LESSON PREPARATION</th>
<th>Yes</th>
<th>Partially</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did the teacher show evidence of having prepared for this lesson before starting to teach it?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For example: Did the teacher seem familiar with the lesson plan, write lesson notes, prepare the necessary materials (like letter acres, word cards, sentence poster, or realia for vocabulary words) and/or write lines, letters or text on the board ahead of time?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Did the teacher follow the lesson plan?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Some of the paper-based binary checklists used for monitoring included additional space for the observer to record comments on what they observed akin to the coaching example above (see Figure 1). The format of items in other monitoring instruments included using a scale to indicate the frequency of teachers' application of specific instructional practices (i.e., very often, somewhat often, not at all).

The format of items in instruments used for research and evaluation varied and included binary checklists, evaluative rating scales, and time-interval sampling methodology such as those used in Stallings or “timed” instruments. See Figure 4 for an example of a time-interval sampling format from the Time to Learn\textsuperscript{52} intervention implemented in Zambia.

**Figure 4. Time interval sampling format in evaluations**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 2A 1. ORIENTATION TO PRINT (top to bottom, left to right, return)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the Zambia time-interval sampling example, the observers record on a tablet what is happening every three minutes. Three-minute time intervals are represented by P1, P2, P3, etc. During each interval, the observer records the behaviors that are observed and the tablet keeps track of time and automatically moves to the next column after each 3-minute interval.

Guidance on identifying an item and response format appropriate for the purpose, observer capacity and other contextual considerations can be found in Item and response formats in the section Understanding Classroom Observation.

Instrument templates provided in Annexes B-F provide examples of potential formats for different purposes.

**Focus of Instrument**

Based on results of the EGR survey, the focus of observation instruments varied across and within all purposes and uses. Most instruments focused to some extent on teachers' use of evidence-based practices related to general pedagogical practices and literacy instruction. While coaching and monitoring instruments focused mostly on teacher behavior, some monitoring and evaluation instruments included items to capture student engagement and teacher-student interactions. This was more common in instruments used to measure quality of instruction. Most instruments recorded student attendance, especially instruments used for monitoring purposes.

Instruments used to support teacher growth through feedback and coaching were aligned with intervention-specific competencies or instructional strategies introduced in the professional development intervention. The instruments reviewed also provided space for coaches to document qualitative evidence or take notes about each behavior observed.

In addition to the content focus of the observation, the instruments included other sections for them to analyze and summarize teachers' strengths and areas for improvement based on the observation. EGR instruments also commonly included a section for an “action plan” for teacher improvement.

To accompany the instrument, most instruments also included a protocol either as part of the instrument or included with coaching resources that illustrated the steps coaches should follow when conducting classroom observation visits and giving feedback to teachers. See Figure 5 for an example of an observation summary section from the Ghana Partnership for Education: Learning program.53

**Figure 5. Analysis section of a coaching instrument**

<table>
<thead>
<tr>
<th>Part D. Analysis and Strategy. Coach completes after each visit and before the post-conference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Sources</strong></td>
</tr>
<tr>
<td><strong>Strengths</strong> (Areas where the teacher did very well)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Areas of Improvement</strong> (Areas where you think the teacher needs more support to do better)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Notes for post-conference visit.</strong> What strategies will you (the coach) employ to help the teacher to improve in areas where he/she is struggling?</td>
</tr>
</tbody>
</table>

Instruments used for monitoring purposes were designed to align closely with teachers’ guides and lesson plans. This indicates programs are appropriately aligning their instruments to what teachers are expected to do. Some targeted very specific behaviors relevant to a specific lesson plan for a given day (e.g., “Did the teacher teach the song?”) while other instrument items were more general (e.g., “Teachers used the I do, we do, you do” approach). Several monitoring instruments also recorded the availability of teaching and learning materials, student attendance in class on the day of the observation, GPS location of schools and other monitoring data, information that may be useful to some programs but is not necessarily required depending on the observation purpose and other considerations. Monitoring instruments that were paper-based included a protocol for observers to provide feedback, either as part of the instrument itself or as a separate document that accompanied the instrument. As discussed in the guidance section, such protocols can be helpful to observers and support reliable data collection.

Attaining consistency and quality in coach-provided feedback to teachers can be a challenge, especially for programs implementing at a large scale and with large numbers of coaches. One way that many programs have addressed this challenge is by using electronic monitoring instruments (e.g., observer completes observation on a smart phone or tablet) which contain a built-in algorithm that automatically generates coaching feedback based on the observation they

---

record. For example, in the Unlock Literacy Measuring Evidence of Quality Achieved (MEQA) initiative, the observation instrument had a pre-programmed analysis, so that the teacher’s “strengths,” “challenges” and “areas for growth” would instantly pop up at the end of the tool. See Figure 6 for another example from Tangerine:Tutor, used on the Kenya Tusome (2015-2019) initiative.

**Figure 6. Automatically generated feedback in Tangerine:Tutor**

While automatically generated feedback can confer several advantages, limitations to its use exist as well. For example, limited evidence exists regarding its effectiveness and whether it makes a difference in the quality and effectiveness of coach feedback. Also, some programs have found it difficult to develop algorithms that produce the appropriate type and amount of feedback for a specific coaching session. Another limitation is that having automatically generated feedback can mislead policy makers and implementers to think that coaches do not need in-depth training on how to be a coach, since they can just get the talking points from their device, which can be problematic if devices do not recognize all scenarios or occasionally give inappropriate guidance.

For observations used to evaluate impact, two different instruments were commonly used to measure the program’s impact on: (1) teachers’ use of targeted instructional strategies, and (2) teachers’ use of instructional time. The evaluation instruments used to evaluate the program’s impact on teachers’ use of targeted instructional strategies were similar in structure and content to instruments the programs used to monitor teachers’ adherence to lessons plans, or fidelity of implementation. Such instruments aimed to measure changes in teachers’ use of literacy instructional strategies compared to baseline and end line and thus focused on whether behaviors were present or not, and in some cases, how often. These instruments were informed by protocols used by external evaluators in other EGR initiatives and adapted to be more specific to the intervention being evaluated.

Evaluation instruments used to measure teachers’ use of instructional time were referred to as “classroom snapshot” instruments. Classroom snapshot instruments use a time-interval sampling item and response format to capture how much time a teacher or students spend on specific content and strategies. Other programs used instruments that had similar time-interval sampling formats but captured activities every three minutes, and the content focused explicitly on

---

reading instruction strategies promoted by the program. The difference between the two instruments is one measures the amount of time teachers spend on general instructional strategies using valid measures of quality instruction such as time on task and ability to keep students engaged (e.g., Stallings Classroom Snapshot); the other instrument is tailored to a specific EGR program and uses a different time-interval sampling format to measure teachers’ use of instructional time on program-specific reading strategies and content.

See Example from practice 6 for more details on how Ghana is using two different observation instruments as part of an impact evaluation to measure a program’s impact on two different aspects: 1) teachers’ use of instructional strategies, and 2) teachers’ use of instructional time.

**EXAMPLE FROM PRACTICE 6**

**Classroom Observation for Impact Evaluation**

To evaluate the Ghana Partnership for Education: Learning, a reading initiative implemented by FHI 360 and partners from 2014-2019, two different observation instruments are being used: (1) a reading program-specific observation instrument referred to as the Structured Observation Protocol and (2) the Stallings Classroom Snapshot instrument, a standardized observation instrument that has been verified as a reliable and valid instrument that can be used across contexts to collect comparable data concerning classroom instruction. The Structured Protocol uses a rating scale of 1-3 to measure the impact of the program on teachers’ use of program-specific reading strategies. The Stallings instrument uses a time-interval sampling format to measure teachers’ use of instructional time, use of different teaching and learning activities, and ability to keep students engaged. The use of the Stallings instrument was important to be able to collect credible data that could be compared between treatment and comparison schools to measure changes in classroom instruction that could be directly attributed to the initiative (Social Impact, 2018).

**Instructional Categories and Behaviors**

The instructional categories included in the instruments that have been used by EGR programs to date, and that were reviewed to inform the development of this resource, varied depending on purpose as well as what programs prioritized as effective teaching practices. Overall, most of the instruments included evidence-based instructional categories and items specific to the reading program, which is the desired approach. The most common instructional categories across all instruments, regardless of purpose, included: early grade reading instructional strategies; checking for understanding; general lesson delivery strategies; supportive and inclusive learning environments; physical classroom environment; lesson preparation; use of instructional materials; and use of objectives. Several programs noted that they periodically reviewed and revised their instruments to items linked with positive outcomes on EGRA results.

Instructional categories of instruments used with coaching primarily included general lesson delivery strategies, EGR strategies, classroom management and checking for understanding. Instruments used for monitoring purposes tended to focus primarily on the topics of lesson preparation, EGR strategies and checking for understanding. Research and evaluation instruments reflected a wider balance across observation categories, including a greater focus than the others on time on task, use of materials, checking for understanding and student engagement.

Within each instructional category, most instruments included observable behaviors that aligned with the instructional category they intended to measure. However, there were several that did not. Such observation data would provide an inaccurate understanding of what teachers were able to do or not do. For example, one instrument had an instructional category referred to as “child-led activities,” but the observable behaviors were behaviors around teaching and learning materials being available and sustainable. The availability of teaching and learning materials may contribute to child-led activities but would not be a proper measure of child-led activities happening. Many instruments included items that were not observable within the classroom observation period. This was more common in instruments used for monitoring purposes, which gathered additional information either for program monitoring information needs or for the MOE. For example, several instruments included questions about the teacher’s background and fluency, existence of a school
feeding program, and existence of water; sanitation and hygiene (WASH) facilities. These items are outside the scope of a classroom lesson on teaching and learning and require further inquiry and observation outside of the class period. While such information might be useful to inform what factors most contribute to learning outcomes, they are likely better suited to another observation instrument on school climate or environment.

**Feasibility and Usability**

Respondents to the GRN survey about classroom observation provided useful insight into many important aspects related to instrument feasibility and usability in the context of EGR programs. Several survey respondents recognized that the feasibility of their instruments needed to be improved. For example, some survey respondents noted their instruments were too long, making them challenging for observers to reliably record information for every item during the course of one lesson. The review of instruments found similar issues with feasibility. For example, one instrument contained more than 80 items, to be observed during one lesson. Some instruments were too text heavy, meaning an observer would likely find it difficult or time consuming to read all of the information provided. Such feasibility issues can be identified and adjusted before data collection if the instrument is properly piloted. Example from practice 7 describes how one initiative in Mali mitigated problems associated with a long instrument by dividing the content into seven separate instruments. This adjustment allowed observers to focus on a specific teaching technique in each lesson. Many EGR programs, particularly those implemented at scale, reporting that that developing and using a separate instrument for both coaching and M&E purposes was not feasible given limited funds, time, and observers. As a result, several programs used the same instrument for both purposes. Coaches used one instrument to observe teachers and provide

**EXAMPLE FROM PRACTICE 7**

Feasibility and Usability Considerations

The Mali Selective Integrated Reading Activity (SIRA), implemented by EDC from 2016-2021, found that one instrument alone could not gather enough information to assist the program in understanding what instructional strategies teachers were able to implement effectively. Therefore, the project established an observation system that divided what had been a long observation instrument into seven separate instruments. Each instrument featured a specific teaching technique for pedagogical coaches to focus on during one lesson (e.g., teaching of decodable words, guided writing). The program found that “chunking” the observation instrument to focus on one specific teaching technique made it easier and more manageable for coaches to collect information in a lesson. Additionally, the information the coaches collected provided more useful insight into the teachers’ skills in each area (GRN survey response, 2018).

feedback, and the same data was used to monitor program implementation.

The review of observation instruments also found that many instruments could be improved to be more user-friendly for the observer. Some of the major issues with usability include:

- Many instruments were too long (some up to 14 pages), which can be difficult for observers to manage as it requires them to flip between pages (or, in the case of electronic instruments, to navigate between screens) to view items, which can distract their attention away from teachers and learners.
- Some instruments written in English used jargon or terminology specific to American English that might be difficult for a second language English speaker to understand. For example, an item on one instrument read: “the teacher provides ‘rich and meaningful lessons’ in language development.” In this instance, rich and meaningful, if translated directly, could be misinterpreted.
- Some instruments had inconsistent and complicated item and response formats requiring the observer to learn how to score using several different formats.
- Some instruments used highly subjective evaluative items without including clear scoring criteria to guide the observer, requiring observers to have more technical knowledge in order to score consistently.
• Some instruments used a checklist format intended to measure the adherence of certain behaviors or frequency, but the rating scale asked the observer to monitor quality (poor, good, excellent) rather than adherence (yes, partially, no) or frequency (never, sometimes, always). Quality rating scales require a more experienced observer and clear scoring criteria.

Guidance on Instrument Features and Content

When selecting or developing an appropriate observation instrument for the intended purpose(s), consider the following questions during instrument review:

• **Is the instrument VALID? How do you know?** Validity is one of the most important instrument features to consider when developing or selecting observation instruments. Instrument validity is the extent to which the instrument measures what it is supposed to measure. In other words, validity refers to the extent to which the instructional categories and observable behaviors on the instrument are informed by evidence to improve positive student learning outcomes. When developing or modifying an existing instrument, check for documented evidence on effective classroom and reading instruction and involve technical and subject-matter experts (e.g., reading instruction or SEL if relevant), as well as experts on educational measurement, to verify the instrument is valid and measures what it needs to measure.

• **Is the instrument RELIABLE? How do you know?** Reliability is another important instrument feature to consider when developing or modifying an instrument. Reliability is the extent to which the instrument consistently measures what it intends to measure. To what extent does the instrument produce consistent scores across multiple observations and when used by multiple observers on the same lesson? Reliable instruments include precise scoring criteria and clear instructions for observers. An instrument’s reliability is determined by several levels of testing conducted by experts in educational measurement. Once a reliable instrument is selected, it still must be tested with observers to assess whether it is reliable across observers.

• **Does the FOCUS of the instrument align with what it aims to measure?** An instrument’s focus will depend on the purpose of the observation and what the instrument aims to measure. Specify first the focus of the instrument to understand what type of information is needed. This information will help inform other decisions about the instrument, including the structure and the most appropriate item and response formats. See Table 3 for examples of instrument focus organized by purpose. For EGR programs implemented in contexts affected by crisis and conflict, instruments might include a focus on social and emotional learning strategies or content concerning safe learning environments. See INEE’s *Psychosocial Support and Emotional Learning in Emergency Settings* and ECCN’s *Safer Learning Environments Assessment Toolkit*.

• **Is the ITEM and RESPONSE format appropriate?** When considering item and response format, consider the following: the purpose of the observation; what information is needed and how it will be used; the expertise and skills of those developing the instrument; and the skills of the observer. Developing or using an instrument with high-inference statements or questions, or items with evaluative rating scales, is harder and requires more expertise on the part of both the individuals developing the instrument as well as the observers using it. Thus, it may require more time and resources to develop an instrument that monitors or measures quality and to sufficiently train observers to be able to assess quality. However, high-inference instruments, when implemented reliably, result in more nuanced data that allow users to better identify gaps in quality classroom instruction. Table 4 provides suggested item and response formats for use in EGR programs. Examples of low-inference instruments for different purposes common in EGR interventions are available in Annexes B, C, D, E, and F at the end of the toolkit.

• **Do the INSTRUCTIONAL CATEGORIES and BEHAVIORS align with the purpose and the instructional strategies that the program is interested in observing?** Instructional categories and behaviors MUST be informed by evidence-based practice and should align with the overall purpose and aim of what the program wants to measure. It is critical that the instrument reflects standards for high-quality and effective reading instruction even in contexts where expectations for teachers may not be as high given their existing knowledge and skills at the time of the intervention. For instruments used to evaluate an intervention, the content should also be aligned with the objectives and expected outcomes. To attend to these two issues,
programs should involve in the instrument development/adaptation process those who are familiar with the program as well as instructional and content-specific experts familiar with the current evidence base for high-quality classroom instruction relevant to the context and specific professional development program (e.g., early grade reading instruction, structured pedagogy\textsuperscript{57} equitable and inclusive instruction\textsuperscript{58}, social and emotional learning practices\textsuperscript{59}). Experts in educational measurement and survey design should also be involved in the development of the instructional categories and behaviors.

- **Is the instrument FEASIBLE to implement?** When selecting or developing an instrument, always keep in mind the skills and capacity of the observers as well as the costs and time required to administer the instrument. An instrument that is too lengthy may not be feasible to implement within the period of a single lesson observation. An instrument with evaluative rating criteria may be too difficult for observers with limited experience in pedagogy and may require additional training, time and costs.

- **Is the instrument USER-FRIENDLY?** When checking an instrument’s usability, consider the format, font, language and sentence structure of the instrument. The format of the instrument should be clear and simple with font and spacing that is easy to read. Avoid having too much text on one page. Be consistent with item and response formats as much as possible to make it easier for the observer to score. For example, switching formats from binary response in one section, to a rating scale of 1-3 in another section, and a different rating scale in another make it complicated for the observer to learn and use the instrument. Keep it simple and consistent. When checking language and sentence structure, consider the language of the instrument and the native language of the observers.

\textsuperscript{57} Young-Suk Grace Kim, & Marcia Davidson, *Promoting successful literacy acquisition through structured pedagogy* (Chevy Chase, MD: URC, 2019).

\textsuperscript{58} For more information on promoting literacy with students with disabilities, refer to USAID’s *Universal Design for Learning to Help All Children Read* (2019). For more information on enhancing girls’ education and gender equality in the classroom, refer to the INEE Guidance Notes on Gender (2019) and relevant sections of Department for International Development (DFID)’s *Girls’ Education and Gender Equality: Education Rigorous Literature Review* (2014).

### Table 3. Focus of observation instruments in EGR initiatives, by purpose

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Instrument focus</th>
</tr>
</thead>
</table>
| **Identify teacher professional development needs** | • Collects information on existing teaching practices and classroom conditions of teachers in a specific context to identify areas where teachers could benefit from a professional development intervention  
  • Focus varies depending on context and specific interests of the needs assessment                                                        |
| **Support individual teacher professional growth and learning** | • Aligns with teacher behavior and instructional practices introduced or promoted in professional development  
  • Collects evidence/examples of teachers using strategies introduced in professional development to guide reflection and feedback on an individual teacher’s strength and areas for improvement |
| **Monitor implementation**                        | • Aligns with teacher behavior and instructional practices introduced or promoted in professional development  
  • Collects information to monitor one or more of the following:  
    • Adherence (presence of specific behaviors)  
    • Frequency (how often)  
    • Use of instructional time (allocation of time to specific content and strategies)  
    • Quality (how well)  
    • Other information needed for tracking fidelity (i.e., availability of materials, attendance, etc.) |
| **Evaluate impact**                               | • Collects information to evaluate impact of teacher professional development interventions on teacher performance  
  • Focus varies depending on how an intervention defines impact of program on teacher performance  
  • If impact is % increase in number of teachers who improved use or frequency of program-specific EGR instructional strategies, use of instructional time, or quality, a program-specific instrument may be used  
  • If impact is determined by a more general measure of quality of instruction, programs should use a reliable and validated instrument for measuring quality. For example, if programs define quality by time on task, they might use the Stallings Classroom Snapshot or if they want to measure both quality and time on task, they might use World Bank’s TEACH observation instrument |
| **Generate new knowledge about teaching and learning processes** | • Collects information to better understand the learning processes and behaviors in classrooms in a specific context  
  • Focus varies depending on the purpose of the research and what specific processes it aims to explore (e.g., student-teacher interactions or teacher read-aloud practices) |
<table>
<thead>
<tr>
<th>Purpose</th>
<th>Item and response formats</th>
</tr>
</thead>
</table>
| **Identify teacher professional development needs** | • Low-inference items  
• Simple binary checklist format (yes/no) or objective rating scale (e.g., evident, partially evident, not evident) checking the extent to which behaviors are observed.  
• In situations where observers have more expertise, programs may choose to use high-inference items that require an evaluative judgement. In this instance, time-interval sampling or simple evaluative rating scales can be used.  
See Annex B for additional guidance and an illustrative template on instruments to identify teacher professional development needs prior to the design of an intervention. |
| **Support individual teacher professional growth and learning** | • Low-inference items  
• Simple binary checklist format (yes/no) or objective rating scale (e.g., evident, partially evident, not evident) with space for observers to document evidence or examples from practice; it is not typically recommended that instruments for this purpose use formats that require observers to make an evaluative judgement about teaching performance. This depends however, on the expertise of the observer and the role the observer has in supporting teachers.  
• Space for identifying strengths and areas of improvement to inform structured feedback and ongoing support  
• Space to record improvement plan of action discussed with teacher and supports needed  
See Annex C for additional guidance and an illustrative template on instruments to support individual teacher growth. |
| **Monitor implementation** | • Low-inference or high-inference items depending on what the instrument needs to measure, observer capacity, scoring criteria, and training  
• If monitoring FOI, binary checklist format (yes/no) or objective rating scale (e.g., evident, partially evident, not evident) of program-specific content could be used; if monitoring use of instructional time, a time-interval sampling format could be used; if monitoring the quality of teacher use of strategies, simple rating scales could be used.  
• Content items should be clear and observable and not require observers to have significant experience in pedagogical background  
See Annex D for an illustrative example of an instrument that measures teachers’ adherence to specific strategies to monitor fidelity of implementation. |
| **Evaluate impact** | • Low-inference or high-inference statements depending on what the instrument needs to measure, the observer, scoring criteria, and training  
• If evaluating FOI, binary checklists of program-specific content could be used (similar to a monitoring instrument); if evaluating use of instructional time, a time-interval sampling format could be used; if evaluating the quality of teacher use of program-specific strategies, simple rating scales could be used  
• If measuring the quality of teacher practices against comparable measures, the item and response format will be determined by the standardized instrument selected  
See Annex E for an illustrative example of an instrument that measures the impact of the professional development on teachers' use of instructional time. |
| **Generate new knowledge about teaching and learning processes** | • Item and response format for observations used for research will depend on the purpose of the observation, the expertise of the observers, and what the instrument aims to measure |
### Instrument Development and Adaptation

**KEY TAKEAWAYS** on Instrument Development and Adaptation

- High-quality instrument development and adaptation is an iterative process that requires collaboration and coordination of multiple stakeholders and experts throughout the entire planning and implementation process.

- Instrument design must always consider key criteria including observer capacity, instrument reliability and validity, instrument focus and information needs, item and response formats, instructional categories and behaviors, and feasibility and usability.

- Piloting an instrument is a critical step to verifying instrument validity and reliability, as well as gathering feedback from observers on their experiences using the instrument.

- A protocol or guide describing how to use the instrument, score the instrument, conduct observation visits, and use the information to give feedback (if relevant to the role of the observer) must be developed to accompany each instrument.

- Observers must be trained to help maximize correct use of the instrument; providing multiple opportunities for observers to practice using the instrument and be assessed on their scoring.

### Experiences from EGR Programs

The process for instrument development and adaptation varied across EGR programs and across purposes. According to survey responses, most programs selected existing instruments used in other programs either within their organization or from other organizations. Development and adaptation of coaching and monitoring instruments was conducted mostly at the country-level and led by technical personnel, and in collaboration with M&E personnel in the case of instruments used for monitoring. Instrument development for evaluation purposes was commonly led by global teams of researchers, evaluators and, in some cases, program managers with feedback from program personnel.

Instrument piloting experiences varied widely, indicating a need to dedicate more time and attention to this critical aspect of instrument development. Piloting was less common in the development of coaching instruments and more common for instruments used for monitoring and evaluation. However, several survey respondents noted they were not able to pilot instruments due to limited time. Donor requirements and time constraints were both reported as significant barriers to having enough time to develop and pilot instruments. This constraint had long-term impact on some programs using instruments for evaluation. In some cases, a lack of piloting before baseline data collection led to use of instruments that were inappropriate for measurement needs. In these cases, evaluators had to continue using the same instrument throughout the life of the program to gather comparable data overtime.

Many respondents noted that one successful aspect of instrument development was the collaborative nature of the development process. One survey respondent noted that including diverse stakeholders (e.g., technical personnel, M&E personnel, education/reading specialists, MOE supervisors, coaches, teachers) in the instrument development process contributed to the development of a quality instrument and garnered ownership over and support for the instrument itself. A few respondents specifically mentioned the involvement of both technical personnel and M&E personnel as important to development of monitoring and evaluation instruments. Other successes mentioned included the following: creating an instrument that was appropriate for observers’ skill level; using field data to revise and improve the instrument; and finding solutions to various challenges the program faced related to classroom observations. See **Example from practice 8** for more information on how the Pakistan Reading Project and the Unlock Literacy MEQA program engaged in a collaborative and iterative instrument adaptation process that used feedback and piloting data to refine their instruments over time.
Guidance on Instrument Development and Adaptation

This section identifies key aspects of the instrument development and adaptation process to consider and plan for in the early stages of program design. The guidance is informed by EGR program experiences to date and the literature on effective classroom observation.

It’s important to note that instrument development and adaptation are not linear processes. Therefore, the structure of this section is not a suggested sequence of activities but rather a summary of key aspects of the process. These recommendations are as follows:

- **Engage a diverse group of stakeholders and experts.** To verify that an instrument is valid and reliable and relevant to the local context, instrument development requires technical experts knowledgeable of the most current research and best practices in early grade reading instruction in LMIC contexts, teacher professional development, equitable and inclusive education, research design, and instrument development and use for classroom observations for the appropriate purpose. These experts may include ministry of education officials at all levels; technical specialists that have expertise in areas relevant to the context and intervention, including early grade reading instruction, professional development, classroom observation, equitable and inclusive education, safe learning environments, and social and emotional learning; lecturers; context experts; monitoring and evaluation personnel; survey design specialists; evaluation design experts; and data analysts. Engaging relevant education officials, local education authorities, observers, and teachers throughout ALL phases of the instrument development and administration process is critical. For some instruments, teachers should also be involved, particularly on instruments used for their own professional development, so that they have ownership over the content and a good understanding of the criteria used for observation. See Table 5 for a list of potential stakeholders to involve and their roles in the development and use of classroom observations for different purposes.
Table 5. Stakeholders and experts to include in classroom observation instrument development, by purpose

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Stakeholders and experts</th>
</tr>
</thead>
</table>
| Identify teacher professional development needs                        | Technical experts (e.g., in early grade classroom and reading instruction, equitable and inclusive teaching practices, safe learning environments, and social and emotional learning)  
|                                                                        | Relevant personnel (e.g., from teacher training institutions and program)  
|                                                                        | Small sample of observers (ideally with previous experience, to inform content and to pilot instrument)  
|                                                                        | Translators  
|                                                                        | Other stakeholders as relevant                                                                                                                                  |
| Support individual teacher professional growth and learning            | Technical experts (e.g., in early grade classroom and reading instruction, equitable and inclusive teaching practices, safe learning environments, and social and emotional learning)  
|                                                                        | Relevant personnel (e.g., from teacher training institutions and MoE) responsible for providing and improving professional development and support to teachers  
|                                                                        | Pedagogical advisors and select head teachers, peer teachers or coaches who are trained to use the instrument and to provide on-going support to teachers  
|                                                                        | Small sample of teachers of early grade reading (select)  
|                                                                        | Small sample of observers (ideally with previous experience, to inform content and to pilot instrument)  
|                                                                        | Translators  
|                                                                        | Other local stakeholders as relevant                                                                                                                             |
| Monitor implementation                                                 | Technical experts (e.g., early grade classroom and reading instruction, equitable and inclusive teaching, safe learning environments, and social and emotional learning)  
|                                                                        | Monitoring and evaluation specialists/personnel  
|                                                                        | Ministry of education officials responsible for the monitoring and supervision of the schools and teachers and quality assurance  
|                                                                        | Small sample of observers (ideally with previous experience, to inform content and to pilot instrument)  
|                                                                        | Translators  
|                                                                        | Other local stakeholders as relevant                                                                                                                             |
| Evaluate the impact of professional development                        | Technical experts (e.g., research, education measurement, primary education/reading, etc.)  
|                                                                        | Monitoring and evaluation specialists/relevant personnel  
|                                                                        | Small sample of observers (ideally with previous experience, to inform content and to pilot instrument)  
|                                                                        | Small sample of observers  
|                                                                        | Translators  
|                                                                        | Other local stakeholders                                                                                                                                           |
| Generate new knowledge about teaching and learning processes           | Lead research team members  
|                                                                        | Technical experts (e.g., in early grade classroom and reading instruction, equitable and inclusive teaching, safe learning environments, and social and emotional learning)  
|                                                                        | Personnel from relevant ministry of education units or teacher training institutions (e.g. individuals responsible for specific areas for which research is being conducted, ethics approval and those with interested in the research).  
|                                                                        | Relevant personnel responsible for providing and improving professional development and support to teachers  
|                                                                        | Translators  
|                                                                        | Other local stakeholders as relevant                                                                                                                             |

- **Select an existing instrument or develop a new one.** Programs will need to decide whether to develop a new instrument or select an existing instrument. Selecting an existing instrument that aligns with the program content and adapting it, rather than developing a new instrument, is highly recommended.\(^6\) If an existing instrument is chosen for adaptation, try to build on one that has already been used in a similar context. If a program wishes to develop a new instrument, seek guidance and consult with experts in educational

---


32 Classroom Observation Toolkit
measurement. When developing or selecting an instrument for evaluation purposes, programs need to take time to select the right type of instrument before using it in the baseline. Impact evaluations require the same instrument for both baseline and endline data collection to show change over time vis-à-vis established indicators. Therefore, the content of the instrument cannot be changed after the baseline data collection. If EGR programs need to develop or select one instrument to serve multiple purposes given the context or limited time and resources, select or develop instruments that reflect the criteria needed for each purpose in terms of instrument focus, item and response format, and format of the instrument. If instruments are not currently available that match the content and context of the program, observation instruments and documentation of their use in LMIC contexts might be available on individual organization or program websites, or in program reports posted in the Development Experience Clearinghouse (DEC). USAID’s Education Links website also contains a repository of EGR program reports and instruments.

- **Review instrument features.** As discussed previously, review the instrument to verify it meets the criteria specific to the intended purpose. Is it reliable and valid? Is the item and response format appropriate to the purpose and capacity of the observers? Do the instructional categories and observable behaviors align with the purpose and focus of the program? Is the instrument user-friendly and feasible to implement?

- **Gather information about previous use of the tool.** Even if an instrument meets the specific criteria, this does not necessarily mean it will be appropriate for a given program and context. To find out, learn how the specific instrument has previously been used. This may require seeking and reading documentation about its use or speaking with individuals who have used the instrument. For monitoring, research and evaluation purposes, reports should be available and normally include information about IRR scores, observer training, and limitations of the instrument or data gathered.

- **Modify content and structure if necessary.** If instruments selected do not meet all the criteria described in this toolkit, modifications may be necessary to improve the instrument, to collect the type of information needed and to adapt it to align with country, programmatic, and educational contexts. However, when using reliable and valid instruments for measuring quality of a program or quality of instruction, avoid making too many modifications to the core features of the instrument, including instructional categories and item and response format. Modifications can reduce an instrument’s reliability and validity, as well as feasibility and usability if modifications do not follow best practices in instrument development.

- **Adapt instrument content to program and country context.** The process of adaptation may vary across programs depending on the purpose of the instrument, its origin, the medium of the instrument and the type and degree of adaptation required. Adaptation commonly involves aligning the instrument with the national curriculum or program-specific approach, the educational context, the language of instruction, and observers’ skill level. Participation of program personnel, local stakeholders and translators during the adaptation phase of the instrument is essential. When revising content of the instrument or translating into another language, translators and relevant technical personnel work together to ensure the translations of specific terminology and concepts are accurate and do not change the meaning. For example, an instrument originally written in English may use the word “print-rich” when describing the classroom environment. If the translator is unfamiliar with this terminology and translates directly into another language, the meaning could potentially change. Even if the instrument does not need translating, a term such as “print-rich” might need to be changed to one more common and relevant to the context. Stakeholders and experts should agree on the most appropriate terms to use in the relevant language, and observers should be trained on their meaning.

- **Pre-test draft instrument.** After instrument adaptation, those responsible for developing the instrument should pre-test the draft instrument in a handful of schools. Pre-testing does not require a cohort of observers to be formally trained, as is the case with formal pilot testing. Rather, it provides those who are developing the instrument with an opportunity to get initial feedback on various aspects of the instrument, including wording of specific items, appropriateness of items given the content, and instrument length. Pre-testing informs the development of the instrument while in progress and can take place more than once if needed. Revisions to the draft instrument can then be made based on the pre-testing.

---


62 The USAID Education Links website is available at [https://www.edu-links.org/](https://www.edu-links.org/).

63 Jilliam N. Joe et al., _Foundations of Observation_.

---
• **Pilot instrument for reliability and validity.** To verify whether the instrument is valid and reliable, an instrument must be piloted. Piloting should take place in a similar context, or in a similar set of schools, as the full data collection. A small number of observers can be recruited for the pilot testing. Consult with technical experts to identify the best method for measuring the instrument’s reliability, feasibility and usability. Piloting is also useful to identify how much time observers need to complete the instrument, to assess if they understand the meaning of each item and how to record responses, to review the quality and consistency of observer responses, and to gather additional feedback about observers’ experiences using the instrument. Piloting also provides an opportunity to understand if the data collection method is appropriate and resulting in the desired information being collected. After piloting, identify and make necessary modifications, along with any needed decisions about modifying observer training.

• **Develop or modify existing protocol for observation visits.** Instruments should be accompanied by a protocol or guide explaining to observers how to appropriately administer the classroom observation instrument. Regardless of the purpose, the protocol should describe the activities observers should undertake before the observation, during the observation, and after the observation, as well as appropriate observer etiquette throughout the process. Scoring guides and criteria should also be included to explain the items that need to be observed and how to score them. Other aspects of the protocol will differ depending on the purpose and the instrument. For example, instruments used by coaches may include guidance on how to give supportive feedback and ask reflective questions. Protocols for research and evaluation instruments may include guidance on gathering informed consent and use of ethical principles.

• **Be open to the possibility of change.** As one GRN survey respondent emphasized, it is important to remain willing to modify an instrument. Instrument modifications will need to occur throughout the adaptation process to verify the instrument is relevant to the purpose, program and context. At any point that the instrument needs modification, review it again against the criteria listed in this resource and pilot the revised instrument to verify its validity and reliability.

Annexes A-F include additional guidance and instrument templates that can be adapted for purposes commonly needed by EGR programs:

• **Annex A** provides a quick reference checklist of recommendations to support the instrument design process.

• **Annex B** is an illustrative template of a low-inference instrument designed to understand teachers’ use of instructional strategies. The format and content could be adapted and used to identify teacher professional development needs prior to developing an EGR intervention.

• **Annex C** is an illustrative template tailored for the purpose of pedagogical coaching. This low-inference instrument measures teachers’ use of instructional strategies. The template provides additional example items to help coaches collect and document information that will be useful during a post-observation visit with a teacher.

• **Annex D** is an illustrative template designed to monitor teachers’ adherence to a specific lesson. An example of a low-inference instrument with a yes/no response format, the instrument’s content is focused on specific behaviors that EGR teachers would be expected to demonstrate during the lesson.

• **Annex E** is an example of a time-interval sampling response format that can help measure teachers’ use of instructional time. The template provided is an example of a low-inference instrument designed to support a specific program’s monitoring and evaluation activities.

• **Annex F** provides a simplified version of a low-inference instrument, using a yes/no response format, that can be used by various individuals (e.g., donor, program or government personnel) to gain a general understanding of how an EGR intervention is being implemented. It is intended to be used for informal classroom visits, not as part of a formal monitoring, evaluation or coaching visit.

Note these templates are illustrative only and should be adapted for different purposes based on the considerations and guidance described in this Toolkit.
Preparation of Observers on Instrument Use

**KEY TAKEAWAYS** on Instrument Development and Adaptation

- Limited capacity of observers to administer observation instruments correctly and reliably is reported as a key challenge in EGR programs. This is due to many factors including poorly designed instruments, limited recruitment pool of applicants, or lack of adequate preparation and training on use of the instrument.
- Training and preparing observers adequately on effective instrument use and reliable scoring is critical.
- Training content should align with the purpose of the observation and observers’ skill level.
- Multiple opportunities to practice using authentic classroom videos and observations in real-life settings are critical components of observer training and will help improve observation quality and inter-rater reliability.
- Once observers are trained and have had multiple opportunities to practice using the instrument, observers should be assessed on their abilities to use the instrument correctly, accurately, and reliably.

Experiences from EGR Programs

Observer Capacity

Observers in current and past EGR programs varied according to the purposes for observation. Many observers for coaching and monitoring purposes came from within the education system. Many were employees of the school: teachers, head teachers, curriculum specialists, or other personnel. Some were local education office personnel who already had a supervisory or support role across several schools. Observers for evaluations, however, were generally individuals hired from local data collection firms or universities. This difference suggests that programs are actively selecting observers to target the right personnel for close-site observation for continuing feedback, and highly objective external observers for formal evaluation purposes, appropriate for each specific context. See the GRN resource *Coaching in early grade reading programs: Evidence, experiences and recommendations* for more information on the role of coaches in observing teachers. Also, see *Planning for Classroom Observation* in this toolkit for more guidance on identifying and recruiting observers.

Many EGR programs noted challenges with observers being able to administer the instrument. For example, some observers found the instruments too long and difficult to complete within the specified amount of time. According to one survey respondent, some coaches “struggled to identify weak practice and struggled even more to provide feedback for improvement.” Another survey respondent stated that observers “struggle to rate measures of teaching quality.” This suggests that observers had limited deep understanding of the content of programs, and limited vision about the quality of teaching that is considered up to program-specific standards. These challenges might be a training and a capacity-building issue but also could be related to the instrument itself. Other challenges related to the observers administering the instrument had to do with operational roles. In one program, coaches were also head teachers. Whenever they went to observe a teacher, they had to leave their class unattended. And some survey respondents reported that observers who were part of the Ministry had difficulties shifting from the role of an inspector or supervisor to a coach. These difficulties suggest that the structures of the observation process were working somewhat outside of the norms of business in the school and system, thus inhibiting the ability to institute an observation process that was operationally sound and sustainable.

---

Observer Trainers

The profile of people who trained observers in current or past EGR programs varied. If the purpose of training was to support coaches’ use of classroom observation instruments, trainers were often technical education experts who collaborated with MOE officials who served similar roles or lecturers from teacher training colleges. When instruments were for monitoring purposes only, M&E personnel led the training of observers. However, for impact evaluations, trainers were usually external researchers or evaluators who worked in collaboration with local data collection firms. According to survey responses, it was uncommon for research or M&E teams to work in collaboration with technical personnel to train observers. This lack of collaboration can be problematic, since knowledge of both reading pedagogy and appropriate use of the tool are equally important for observers to be able to gather reliable data.

Training Content and Approach

A review of EGR program experiences indicates observer training averages approximately five days. Training content is similar across purposes and includes topics such as how to use/fill out the instrument, what to look for when observing, and how to score appropriately. In programs where observers provided on-going support to teachers, additional training content included: how to conduct pre-and post-observation visits; how to give feedback; and where relevant, how to provide on-going individualized support to teachers. In programs where observers collected information with tablets, training also included tablet use. See Example from practice 9 for a brief description of Read Liberia’s approach to training observers.

Providing observers with opportunities to practice was a common approach evident across many programs. Training incorporated watching videos and conducting role-play classroom practice activities to give opportunities for observers to practice observing a lesson, use the instrument, and in some cases, provide feedback. The use of videos during training was common, with many survey respondents highlighting videos as “invaluable” and “essential” to training observers and assessing IRR.

Assessing Observer Capacity

GRN survey responses indicated significant variability in whether programs assessed observers’ ability to reliably record information on the observation instrument. Research and evaluation instruments require assessing observers for inter-rater reliability (IRR) and documented IRR scores in the evaluation reports. Rigorous IRR testing was less common for instruments used for both coaching and monitoring purposes. However, several programs noted they conducted some kind of informal assessment of observers’ use of the instrument, as well as inter-rater reliability. This assessment was commonly conducted through on-site observations, a review of observer-trainees’ instrument responses, and whole group discussions.
Survey respondents described several challenges due to observer capacity. While classroom observation for supervision and monitoring purposes may be generally somewhat familiar within current education systems, observation for promoting teacher growth is generally a new concept for many. In some cases, local education officials who seemed to be best suited for taking on the role of observer did not have enough time or struggled to shift from a supervisory role to a mentoring role. Even with training, teachers who assume a coaching role may struggle to provide quality support due to their own similar instructional challenges. Achieving a sufficiently high level of inter-rater reliability, which is most important in evaluation observations or when measuring quality, was a challenge, although many noted that with extensive training and practice, they were able to achieve good IRR. Two programs noted that it took numerous trainings and several years for observers to consistently produce reliable data.

Several respondents emphasized the importance of videos when assessing observer capacity. For example, one survey response from Columbia University’s Telugu Literacy Enhancement Program (T-LEP) (2015-2019) stated that “videos of best practices help enumerators know more precisely the kinds of practices to look for and how to rate different degrees of teacher performance.” Videos used in training were used mostly for assessing observer IRR on instruments used for monitoring, evaluation and research purposes. **Example from practice 10** describes the approach the Nigeria’s NEI Plus initiative used to train and assess observers for inter-rater reliability.

---

**EXAMPLE FROM PRACTICE 10**

**Practice and Observation During Training**

As part of the evaluation activities of the USAID-supported Nigeria NEI Plus initiative (implemented from 2015-2020 by Creative Associates and partners), observers participate in training before large-scale Early Grade Reading Assessment (EGRA) exercises. The training averages between three to eight days depending on whether new observers are involved or adaptations are made to the instrument. As part of the training, coaches practice using the observation instrument on paper by watching videos of real classroom teaching. After the observation, the group debriefs and discusses scoring to harmonize interpretations and increase reliability. The training also introduces the tablet and provides observers with multiple opportunities to practice and be assessed on the tablet with classroom videos. Observers have two days of practice in schools prior to real data collection, followed by debrief to discuss school practice observations and experiences and provide training on issues of concern based on the feedback from school practice.

When observers are practicing on-site, data collected on tablets is uploaded and reviewed immediately to verify reliability. An IRR test is administered to observers during training, and only those who score 90% and above are selected for actual observation. Based on their performance during the training, approximately half of trainees are identified to serve as observers. During data collection, observers are accompanied to schools by trainers or other qualified technical working group (TWG) members so they can be directly supervised and debrief at the school after data is collected. The data, which is uploaded to a server, is also reviewed for overall quality (GRN survey response, 2018).

---

**Guidance on Observer Preparation**

Adequately training and preparing observers on effective instrument use and scoring is critical. When identifying observers and planning training on instrument use, consider the following:

- **Allow adequate time for training preparation.** Begin recruitment early so that there is enough time for trainers to prepare. Consider training more people than needed in order to retain replacements or substitutes who can be called upon if an observer is not able to successfully complete the training or is unavailable when the time comes for data collection. To garner support and involvement from the government or other relevant authorities, it is imperative to include Ministry of Education personnel and other stakeholders as part of training so that they are aware of the new teaching methodologies and the observation instrument and can properly oversee and provide support.
• **Consider observer profiles and capacity needs.** The prerequisite skills required of observers will depend on the purpose of the observation. Factors to consider for each purpose:

- If the purpose of the observation exercise is to improve teacher professional growth and learning, observers need to be familiar with effective teaching practices and quality learning environments for the grade level in which they are observing.

- If the purpose of the observation exercise is for monitoring implementation, observers need to be familiar with the program. If the program is monitoring fidelity of implementation, priority should be for observers to score reliably, which requires intensive training on the instrument itself and assessing for IRR.

- If the purpose of the observation exercise is to evaluate impact of an intervention or initiative, observers need to be able to score reliably, which requires extensive training and IRR assessment. If instruments use high-inference items to assess quality, observers may need more technical experience than for low-inference items. See Table 6 for a description of common observer profiles and training needs.

• **Align training content and approach with the purpose of the instrument and observer profile.** Observers need to understand the purpose of the instrument and how to appropriately use it. This involves training observers on the content and structure of the instrument, what the instrument measures, what to look for when observing, how to record information accurately, and what protocols to follow for observation visits. If the instrument uses a rating scale format or high-inference questions, observers need more time to understand scoring criteria for each item. Depending on the purpose of the observation, other training content may be needed after observers know how to use the instrument. For example, if the observer will provide feedback after the observation, training content should also include how to collect evidence of teaching practice and identify strengths to inform specific feedback. See Table 6 for common observer profiles and skills required.

• **Include multiple opportunities for modeling and practice.** Training should model expected behaviors of the observer and provide multiple opportunities for the observer to practice. When modeling the use of the instrument, trainers can use a video of a lesson or observe a lesson in-person together and later discuss as a group how observers recorded information until everyone has a common understanding of what each item means and how to appropriately score it. If an observer will give feedback to teachers after the observation, trainers can also model how to use the instrument to identify strengths and needs, and how to give effective feedback in a supportive manner. Trainers can then provide opportunities for observers to practice through pair work, group work or role play.

• **Use videos to support observer practice.** The use of videos of teachers delivering a lesson in an actual classroom is essential to an observation training. Videos help observers understand the teacher and student behaviors included in the instrument and how to score items accurately. They can also be used for assessments of observer-trainees’ reliability in scoring. Videos should be selected or prepared in advance of the training and should be relevant to the specific context as much as possible. Consider recording examples of teachers in the context of demonstrating or not demonstrating practices that are in the classroom observation instrument that observers need to recognize, or find videos that have already been created. Prior to training, trainers need to be familiar with the videos and know when in the videos different behaviors can be observed so they can call trainees’ attention to it. They should complete observation instruments based on videos in advance to have prepared models and samples.

---

Table 6. Observer profiles and skills required, by purpose

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Observer profiles</th>
<th>Skills required</th>
</tr>
</thead>
</table>
| Identify teacher professional development needs | Program personnel | • Fluency in the language of instruction and the instrument  
• Ability to use instrument as intended  
• Ability to accurately and reliably record information  
• If observers will assess quality and score high-inference items, an understanding of effective instructional practices and other behaviors |
| Support individual teacher professional growth and learning | Instructional specialists, coaches, peer teachers, head teachers or coaches or individuals in MOE responsible for providing on-going pedagogical support to teachers through feedback and coaching | • Familiarity with effective teaching practices and quality learning environments for grade level  
• Ability to work with adults as learners  
• Fluency in the language of instruction and the instrument  
• Ability to use instrument as intended  
• Ability to accurately record information  
• Ability to give supportive feedback  
• Ability to use information collected to provide pedagogical support to teachers |
| Monitor implementation | Program and/or MOE personnel | • Familiarity with the specific PD intervention  
• Fluency in the language of instruction and the instrument  
• Ability to use instrument as intended  
• Ability to accurately and reliably record information |
| Evaluate program impact and effectiveness | Enumerators external to MOE/program | • Fluency in the language of instruction and the instrument  
• Ability to use instrument as intended  
• Ability to accurately and reliably record information  
• If observers will assess quality and score high-inference items, an understanding of effective instructional practices and other behaviors |
| Generate new knowledge about teaching and learning processes | Researchers and trained data collectors or enumerators | • Fluency in the language of instruction and the instrument  
• Ability to use instrument as intended  
• Ability to accurately and reliably record information  
• If observers will assess quality and score high-inference items, an understanding of effective instructional practices and other behaviors |

Observers can then watch the videos in the training and practice recording observable behaviors on the instrument. If observers record items differently, the trainer can identify whether the instrument needs adjusting or if the observers need further explanation. If instruments are using a binary response format, videos should provide examples of teachers demonstrating practices listed on the instrument. If instruments are using a rating scale, videos might include examples of low- and high-quality instruction. Use of videos will help to improve levels of observation quality and improve IRR among the observers. Include classroom visits as part of training. Provide opportunities for observers to practice administering instruments in a real school setting. After observers have had opportunity to use the instrument in the classroom, they may have questions that they may not have thought about when practicing how to use it in a training setting.

- **Assess observers’ abilities.** Once observers are familiar with the instrument and have had multiple opportunities to practice, trainers should assess observers’ abilities to use the instrument correctly, accurately, and reliably. Observers should be assessed on the following protocol procedures: recording information correctly and accurately; completing all required sections; scoring items consistently (reliably) across observations; and scoring items consistently across observers (inter-rater reliability). If observers
do not understand the items and do not score them accurately and in the same way, the information they collect will not be useful. Trainers should be experienced and familiar with best practices for assessing observers, particularly for assessing IRR in a feasible manner. EGR programs reported that they used videos to assess IRR in a training environment, as well as in a school setting by reviewing and comparing data collected by pairs of observers. Depending on the purpose of observation, the way in which EGR programs assess observer performance may differ from standard IRR testing typically required when measuring quality as part of research or evaluations. For this, consider the level of rigor required for the observation purpose and the associated time and costs when planning. Consult with experts to provide guidance on the most appropriate procedures and statistical analysis for determining IRR. Results of IRR scores should be documented in reports where relevant.

- **Consider on-going training and support.** Several survey respondents emphasized the need for ongoing training of observers, as each year observers improved in their understanding of the instrument. Follow-up training is helpful for introducing key concepts for new observers, while also reinforcing learning for experienced observers. Follow-up training can be shorter than the initial training, and can focus on honing certain skills, problem-solving challenges, or introducing new items or content if modifications have been made to the instrument.

**Instrument Administration**

### KEY TAKEAWAYS on Instrument Administration

- EGR programs use both paper-based instruments and electronic versions administered on tablets or mobile phones (using an app). In some cases, observers record information on paper and then transfer it to a mobile device to facilitate uploading to a server.
- Several issues must be considered when deciding whether to use a paper instrument, e-version or both. These include the context, costs associated with paper and e-versions, time needed for procurement, and time needed for training on the instrument.
- Determining the frequency and duration of instrument use during the early stages of planning is necessary to accurately budget for observation visits and allow adequate time for instrument development and adaptation.

**Experiences from EGR Programs**

**Instrument Medium (Paper or Electronic)**

EGR programs reported using classroom observation instruments in paper format, electronic format and a combination of the two. In programs that conducted observations for coaching purposes only, instruments were mostly paper-based. Some programs used electronic versions (“e-versions”) of instruments, or versions of the instrument that were provided on tablets or a mobile phone using an app or software. In some cases, both a paper version and an e-version of the same instrument were used. This approach was used due to challenges associated with usability of an e-version of the instrument. In these cases, the observer first recorded information on the paper instrument while observing, then entered that data into the tablet or mobile phone version of the instrument. This dual approach was used to accommodate observers’ need for a user-friendly instrument, while at the same time allowing for the data to be uploaded to a central server to facilitate data collection monitoring and future data analysis. See **Example from practice 11** for a description of how two different initiatives use an electronic observation instrument and tablet software to collect monitoring data.
Most survey respondents valued having an electronic-based system for data collection. Several found the software’s ability to generate automatic feedback based on observation data to be helpful to coaches during their post-lesson teacher discussions. Many reported that applications that enable observations to be administered and recorded on mobile devices were user-friendly. The ability to use the app while offline was also important for those with limited internet access. Some survey respondents noted the importance of being able to have data immediately uploaded to a server for analysis and results displayed onto a digital dashboard for real-time monitoring. Example from practice 12 details how the Kenya Tusome program makes data available on a “digital dashboard.” See also Figure 7 to view a sample dashboard from the Unlock Literacy MEQA program.

Example from practice 11
Electronic Instruments

**Uganda’s School Health and Reading Program (SHRP)**, implemented by RTI International and partners from 2012-2019, uses a tablet-based tool with Tangerine:Tutor software that serves the dual purpose of guiding coaches’ support and collecting monitoring data. The tool is used by SHRP staff as well as various government stakeholders, including coaches, who observe all SHRP-supported teachers. Now in its seventh year, SHRP has collaboratively revised the tool numerous times, based on evidence collected and user feedback. The tool provides scripted feedback that coaches can give to teachers. It also is used to collect program monitoring that is uploaded to a digital dashboard for review (GRN survey response, 2018).

The **Unlock Literacy MEQA project**, implemented by World Vision, is currently piloting an e-monitoring system in 500 schools in three countries. The MEQA project uses a mobile app tool to capture observation data. The digital tool consists of three components: (1) mobile app digital survey; (2) provision of instant feedback for coaching; and (3) online dashboard to show progress on indicators. Collected data is immediately uploaded into a cloud server linked to an online dashboard which presents summarized data visually for the public (GRN survey response, 2018).

Example from practice 12
Digital Dashboards

As part of the USAID-supported **Kenya Tusome** early grade reading activity (implemented by RTI International and partners from 2015-2019), Curriculum Support Officers served as instructional coaches observed and recorded observations with a tablet using the Tangerine:Tutor software developed by RTI International. When coaches had access to the internet, they could upload data to a server, which was then available via a dashboard for coach supervisors and national-level Ministry of Education officials. The observation data dashboard was designed to help education authorities monitor whether coaches are carrying out their duties, and to identify where gaps may exist in terms of supporting specific teachers (Kipp, S., Pouezevara, S. & Piper, B. 2018).

---

66 A dashboard is a digital information system that aggregates uploaded data and displays results through visual representations such as charts and graphs.


---

Classroom Observation Toolkit 41
While most survey respondents agreed that classroom observations that are administered on mobile devices are advantageous in many ways, several respondents highlighted some disadvantages or potential challenges with them. Many preferred observers use paper instruments during the actual classroom observation either because the observers were not comfortable with the electronic version or they felt the presence of an electronic device during observation created a distraction either for the teacher or the students. Others suggested that using a paper-based instrument in the class allowed observers to record what they are observing more easily since the complete instrument can be viewed, as opposed to a tablet- or mobile-phone based instrument that requires observers to scroll for an item, potentially causing delays in their ability to record information. Others noted the significant cost of obtaining and maintaining electronic equipment, and the need for more extensive observer training in the proper use of technological devices and software. Some respondents, however, noted that the cost of hardware and maintenance was offset by improvements in the quality of data captured and the immediate nature of data entry. Moreover, if programs conduct classroom observation multiple times, and are using mobile devices for student assessment, the cost of hardware and maintenance may be worthwhile and may be less than paper-based data collection.

Frequency and Duration of Instrument Use

The frequency and duration of observations varied extensively across EGR programs. Observations for coaching and monitoring purposes were conducted on a regular basis ranging from bi-monthly to monthly or quarterly. Most observers conducted an observation for the full lesson period.

Guidance on Instrument Administration

The recommendations below serve to support EGR programs in the planning and implementation of a classroom observation exercise. Guidance to support quality and effective instrument administration is summarized below:

- **Identify instrument medium (e.g., paper or e-version).** During initial planning, programs need to decide whether they will use a paper-based instrument and/or an electronic version. Issues to consider include the context in which the instrument will be used, as well as costs and benefits associated with paper instruments and e-options. In some contexts, electronic versions may require more time to prepare and potentially require
more time and resources for observer training and data management. However, e-versions may also present certain advantages, such as built-in data quality controls, ability to upload, monitor and use data quickly, and ability to aggregate data collected with various instruments. When deciding which medium is the most appropriate, consider the context and whether it will support e-data collection, instrument format, observer needs and preferences, number of times the observation will be conducted, and the costs and resources needed to support data management and storage.

- **Decide the frequency and duration of instrument use.** During initial planning stages, decide how often observers will conduct an observation with the same teacher, the amount of time required for each observation, and the duration period in which observations will be conducted. For example, if the purpose of classroom observations is to support professional growth and learning of teachers or monitor programs, observers need to conduct observations more frequently than for an impact evaluation, which typically occurs twice, once at the beginning of an intervention and again at the end of the intervention. By determining the frequency and duration of observations in advance, programs can accurately budget for visits and develop an instrument that is appropriate for its planned use. See Table 7 for examples of how frequently classroom observations may be conducted based on their intended purpose. Note that limited evidence exists as to what the ideal frequency and duration of observation visits should be in LMIC contexts. However, research studies in Kenya have demonstrated that the more sustained contact time teachers have with a coach and the more consistent classroom observations can be, the greater impact the coaching relationship can have on teacher behavior and student achievement.⁶⁹ Therefore, if observations are for coaching purposes, the more time a coach is able to spend with a teacher to establish trust and provide on-going support, the better.

### Table 7. Frequency of classroom observation, by purpose

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify teacher professional development needs</td>
<td>Once, prior to a professional development intervention</td>
</tr>
<tr>
<td>Support individual teacher professional growth and learning</td>
<td>Bi-monthly, monthly or other specific number of times per school term</td>
</tr>
<tr>
<td>Monitor implementation</td>
<td>Monthly, quarterly or yearly</td>
</tr>
<tr>
<td>Evaluate program impact</td>
<td>Two times: prior to the start of a program (baseline) and after a program has ended (endline); some programs may also conduct classroom observations a third time (mid-line) to gauge program impact</td>
</tr>
<tr>
<td>Generate new knowledge about teaching and learning processes</td>
<td>Varies depending on the research design</td>
</tr>
</tbody>
</table>

Dissemination and Use of Results

**KEY TAKEAWAYS on Dissemination and Use of Results**

- The dissemination and use of observation results in EGR programs varied depending on the intervention as well the purpose of the observation.
- Results of observations used solely to support individual teacher growth through coaching were used mostly to inform feedback to individual teachers as well as to identify improvement goals and type of follow-up support needed. Coaches compiled results of teacher observations from each school into a summary of results to discuss and share with heads of schools or district-level personnel.
- Observations used for monitoring or monitoring and coaching together were most often conducted in collaboration with MOE officials responsible for monitoring and supervision of schools; aggregated results were shared with local district government officials to illustrate needs to help inform where support to schools was needed.
- EGR programs use results from observations for evaluation purposes to discuss implications for practice and adjust programming where needed. These results are also shared with donors through regular reporting or government counterparts and other stakeholders through presentations, policy briefs, and conferences to inform policymakers.

**Experiences from EGR Programs**

Dissemination and use of results in surveyed programs varied. In programs where observations support teacher growth through coaching, results were primarily used to inform feedback to individual teachers, as well as identify improvement goals and the type and level of support needed. In addition, some programs shared a summary of findings about teacher practices with head teachers or district level personnel to discuss progress and problem-solve instructional challenges together. Other survey respondents stated that information collected by coaches is also shared with program personnel or trainers responsible for training and professional development.

Programs that used the observation for coaching and monitoring, or monitoring only, often did so in collaboration with MOE officials responsible for monitoring and supervision of schools. Several programs mentioned sharing aggregated findings with local district government officials, who use the information about needs to inform their support to schools. See **Example from practice 13** from Mali for more details. One program reported sharing data with pre-service teacher training institutions so that faculty would be informed about what is happening in schools and could adapt their pre-service instruction. Another program, described in **Example from practice 14**, found that data helped program leaders to identify and acknowledge excellent teachers. This acknowledgment provides an important non-monetary incentive for teachers to pursue excellence.

All EGR programs share results from evaluations with donors through regular reporting. Program personnel review findings to discuss implications for practice and adjust programming as needed. Programs may also share findings with government counterparts and other stakeholders through presentations, policy briefs, and conferences—with the intent to inform policymakers or other stakeholders responsible for planning teacher professional development interventions. Results from experimental and quasi-experimental studies are commonly shared through peer reviewed academic journals and academic conferences, 70 organization-sponsored blogs, 71 and stakeholder meetings to discuss implications for future practice.

---


Guidance on Dissemination and Use

During an initial design phase, programs need to decide and plan for how information will be disseminated and used, and if shared publicly, who the audience will be. This will ensure enough time and costs are included in the initial workplan and budget. This section highlights factors to consider when planning for dissemination and use of observation data. Guidance on specific actions to undertake includes the following:

- **Identify the purpose and audience for information sharing.** When deciding on the purpose of dissemination and audience, consider what information might be helpful for others and why. Both the purpose of sharing and the audience will depend on the type of information collected. For example, if observations collect data on an individual teacher’s strengths and needs, or a group of teachers from a school or cluster, this information might be helpful for coaches, head teachers and district-level supervisors to help problem-solve changes or tailor additional support. If observations found that most teachers in a cluster of schools were not consistently giving students enough time for the “you do” approach or independent practice in the lesson, the coach and head teachers can discuss how they can provide additional support to the teachers or the program. If observation data is collected for monitoring purposes, the program will need to develop other protocols to be used in conjunction with the classroom observation instrument to explore the “hows” and “whys” behind the findings. If teachers are not implementing the “you do” part of a gradual release model of instruction, why not? Once this question is answered, program implementers can then adjust the approach to include additional training and support to teachers. To identify the appropriate audience, think about who the stakeholders are that would be interested in observation data. Consider how to involve stakeholders in the review and dissemination of results.

- **Decide how to communicate findings.** Consider the target audience and the best strategy for communicating information. Ways in which findings can be shared include reports, policy briefs, meetings, conferences, articles, blogs, newspaper articles, radio and other media. When deciding how findings will be communicated, identify the specific purpose of sharing information with different groups, what and how much information would be useful, what level of technical detail is appropriate, and how user groups prefer to get information. For example, the data shared with coaches might differ from what is included in a technical impact evaluation report. Decide the most appropriate way to present the information (e.g., tables, graphs or infographics) so that the results are understandable and relevant to the audience. Before disseminating findings broadly, share with a small group. This can help identify what content and formats people find accessible to understand.

- **Identify information sharing supports.** Map out the flow of information—or where and to whom information will be given, how and when—at each stage for data entry, data analysis, and dissemination of findings. In addition, identify what types of human resources and equipment the program will need at each phase. If computer-based systems are required, identify steps to secure electronic equipment and train personnel to appropriately use it.
• Include detailed costs and timeline in workplan. Include dissemination activities in the workplan and budget. Major costs to consider include: human resources; IT equipment and support; and printing if relevant.

Once information is collected, include follow-up work to explore the “hows” and “whys” behind the findings. For example, if teachers are not consistently using an instructional strategy that was promoted in the training, programs can follow up to explore reasons why this is the case.

### Planning for Classroom Observation: Additional Considerations

#### KEY TAKEAWAYS on Planning for Classroom Observation

- Effective instrument development and use require extensive planning, time and resources.
- Understanding the broader educational context and the interests and needs of the government should be the starting point for planning any instrument development and adaptation process.
- Long-term coordination and communication between multiple stakeholders and experts throughout the planning, development, and implementation process is essential.
- Developing a work plan and budget will help facilitate successful implementation.

Effective development and use of observation instruments require extensive planning during initial stages of an initiative. This section highlights additional considerations and guidance for planning a classroom observation exercise. Recommendations are as follows:

- **Understand the educational context.** Improve potential for acceptance and sustainable use of classroom observation instruments beyond the life of the program by developing instruments that are relevant to the education context, and align with the interests and needs of government. When making decisions and planning for the use of observation instruments, review the broader context to build on or improve existing structures rather than duplicate efforts that may not be relevant to the context. Key factors to identify and consider:
  - Country’s education priorities and benchmarks for learning
  - Language of instruction
  - How quality teaching is defined and valued within the country context
  - Current purpose and use of observation instruments (purpose of observation, who observes, type of instrument, how the ministry uses information, etc.)
  - Teachers’ attitudes, perceptions and previous experiences with classroom observation
  - Areas of collaboration with ministry officials and local stakeholders to build on and strengthen existing structures and use of classroom observations
  - Agreement on how both the ministry and the program will use observation data.
  - Other areas as needed given the context

In crisis and conflict-affected contexts, rapid assessments are often conducted to understand the educational context at the time of the emergency or during initial recovery. For more information, see USAID’s *Rapid Education and Risk Analysis (RERA) Toolkit* that helps to identify contextual risks. Also, see INEE’s *Guidance Note on Conflict Sensitive Education* for considerations when conducting a conflict analysis.
- **Facilitate communication and coordination with relevant government actors at national and district levels as well as school heads and teachers.** As discussed previously, essential to effective instrument development and implementation processes that are sustainable and relevant is the long-term coordination and collaboration between multiple stakeholders and experts throughout the entire planning, development and implementation process. This includes consistent communication with, as well as engagement and active participation of, relevant ministry and government officials who may not necessarily be as directly involved in the instrument development process as some of their other MOE colleagues but still should be involved during initial planning stages and kept abreast of key developments throughout the process. Head teachers and teachers should also be well-informed of the purpose of the observations and their level of involvement and assured that the overarching purpose is to support their professional learning and growth and improve classroom instruction. Involving school heads and teachers in initial discussions and throughout planning will help minimize any fear or resistance that can potentially occur if teachers are not accustomed to classroom observations as a norm or have had negative consequences in the past.

- **Identify and recruit observers.** Programs need to determine the appropriate persons to administer the instrument and what extent of training and preparation they need. Observer profile and required skill level will vary depending on the purpose (see *Preparation of observers on instrument use* in section EGR Experiences and Guidance). Across all purposes, observers should be fluent in the language of the instrument and the language of instruction provided by the teacher. Other factors to consider:
  - If the purpose is to improve teacher professional growth and learning, observers may be peer teachers, instructional coaches, head teachers, or pedagogical advisors within the ministry.\(^2\)
  - If the purpose is for monitoring implementation, observers may be monitoring and evaluation personnel or Ministry officials whose role is to monitor and supervise schools (different than those whose roles involve providing support to teachers). To minimize bias, observers should have roles outside of those implementing the program directly.
  - If the purpose is to evaluate impact of an intervention or initiative, observers are often external assessors or enumerators hired from a data firm. Observers should be external to the program and free of any conflict of interest.

- **Develop workplan and timeline.** Start planning early to allow for input from all stakeholders, and time for piloting and revision. Once information needs and purpose are identified, map out key activities, individuals responsible, and develop a workplan and timeline. Time considerations include:
  - Recruiting of observers
  - Selecting and adapting an instrument
  - Translating the instrument
  - Pre-testing the instrument
  - Eliciting feedback
  - Making revisions
  - Obtaining instrument approval from relevant authorities, program staff, and the donor
  - Developing an instrument guide
  - Printing instruments, training documents, etc.
  - Developing or securing video footage of teaching practice
  - Training observers
  - Assessing validity and reliability (if appropriate)

\(^2\) See Marion Fesmire and Amy Pallangyo. “Teacher professional development and coaching in early grade reading programs” (Chevy Chase, MD: Global Reading Network, 2018) [Webinar].
- Conducting observations (consider frequency and duration of instrument use)
- Inputting and analyzing data
- Disseminating information

- **Develop a budget.** Proper instrument development also involves significant costs. Some major costs to include in program budgets are:
  - Instrument development and adaptation
  - Use of technology (i.e., tablets, e-monitoring software, etc.)
  - Instrument translation
  - Pre-testing with small sample of enumerators
  - Observer training (venue, equipment, materials, observer fees if relevant, assessment costs, etc.)
  - Data collection
  - Data analysis and dissemination of findings
  - Other cross-cutting costs (human resources, printing, transportation)

Adhering to the guidance described above will contribute to the development of a quality instrument and a successful classroom observation exercise.
Summary and Future Directions

Evidence from this review of literature and EGR program experiences indicates a critical need for improved classroom observation instrument development and use in reading improvement programs. Developing a quality observation instrument that collects the kind of information a program needs is a complex and time-consuming process that requires extensive planning and collaboration. However, it is feasible with careful consideration and planning at each stage of the process, including: instrument development and adaptation; training and preparation of observers; instrument administration; and dissemination and use of findings.

The Classroom Observation Toolkit for Early Grade Reading Improvement provides basic guidance and considerations as an initial step towards more effective instrument development and use of classroom observations in EGR programs. However, more work is needed. USAID and other donors, governments, and the many organizations and individuals supporting EGR initiatives can continue to advance efforts to improve teaching and learning through use of classroom observation by further discussion and collaboration on the following:

- **Improved planning.** Limited time to develop and properly pilot instruments was a key challenge for many EGR programs, particularly for baseline assessments. EGR programs need to include instrument design in their implementation workplans at the proposal stage and consider the time and budget needed for instrument adaptation and piloting. Donors can help minimize these barriers at the solicitation stage by emphasizing the importance of well-developed observation instruments, allocating adequate time, and requesting programs to prioritize planning for instrument development.

- **Improved quality assurance of instruments.** An important finding from this review was the inconsistency across instruments in terms of quality construction. While many instruments had instructional categories and items informed by best practices in EGR instruction and directly aligned with their reading approach, the other instrument features were limited in quality. Key issues identified included lengthy and text-heavy instruments, inconsistent item and response formats, incomplete phrases, and high-inference questions with limited to no scoring criteria. Programs should always consider the capacity and background of the observer when developing instruments. Programs need to also consult with individuals experienced in survey design and item construction to verify that instrument features and content are appropriate to its purpose and will gather valid information. Finally, prioritizing the assessment of observers during training is important for improving the likelihood that information gathered will be reliable.

- **Expanded focus on processes and quality of teaching and learning.** This review indicates that most EGR programs use classroom observations for the purposes of coaching, program monitoring and evaluation. However, instruments are primarily quantitative and structured to monitor and measure teachers’ adherence to specific instructional strategies as evidence of a program’s impact on teacher performance. These results are analyzed in conjunction with early grade reading assessment results and serve as a proxy measure of quality instruction. While the information is important to understand potential linkages to student achievement, the focus of the classroom observation then becomes on “teachers teaching” rather than the quality of instruction or “students learning.” Few instruments capture how students respond to the lesson, how the teacher and students interact, or how the teacher impacts student learning in each lesson, all key determinants of improved learning outcomes. To advance efforts in providing quality instruction that is more locally contextualized, classroom observations can serve to capture the more complex teaching and learning processes within specific contexts. EGR programs can then corroborate the findings from evaluations to generate a better understanding of teaching and learning processes in different contexts and what might define quality instruction within a given context.

---

• **Increased experience sharing and reflection as a community of practice.** More information-sharing and collaborative discussion within and across EGR programs on instrument development and use will help generate best practices and thus, improve the overall quality of instrument development and use. External evaluators and program teams should create opportunities for more dialogue and sharing on the advantages and disadvantages of using specific instruments in evaluations. Programs developing and using FOI instruments to monitor adherence can benefit from shared discussion and reflection on FOI data and what it is capturing. M&E teams and technical teams should collaborate more on how to make certain that using one instrument to serve both as a coaching tool and a monitoring tool can be done without minimizing the valuable role observation plays in promoting teacher reflection and tailoring support to their individual needs. Furthermore, EGR programs should share their instruments with one another through program reports, submitting them to publicly available websites such as the Development Experience Clearinghouse (DEC), as well as sharing videos and materials used in training.

The overall goal of any classroom observation is to gather information that can be used to help teachers and improve the quality of teaching and learning. Whether information is used to support individual teacher reflection and goal-setting, to monitor and evaluate teacher professional development, or to inform policy and improve programs at scale, all purposes require the use of a well-developed instrument that aligns with its purposes and measures what it needs to measure. As this resource has illustrated, developing and administering a high-quality classroom observation instrument requires intensive planning, expertise, time, budget, and well-trained observers. Furthermore, it is a long-term iterative process that requires the sustained involvement and regular feedback of multiple experts and stakeholders. While perhaps challenging, it is feasible. With proper guidance and support from trained experts, EGR programs can begin to improve their instruments to collect data that will be helpful for the program and, ultimately, the teachers and students.
Bibliography


Kim, Young-Suk G., and Marcia Davidson. Promoting successful literacy acquisition through structured pedagogy: Global Reading Network Critical Topics Series. Chevy Chase, MD: Prepared by University Research Co., LLC. (URC) under the Reading within Reach (REACH) initiative for USAID’s Building Evidence and Supporting Innovation to Improve Primary Grade Assistance for the Office of Education (E3/ED).


Pflepsen, Alison, Karon Harden, and Simon King. “Improving Early Grade Literacy Instruction and Student Outcomes: Findings from a Cluster Randomized Controlled Trial in Northern Nigeria” (in press).


Annex A. Summary of Guidance for Instrument Development and Use

During a classroom observation, observers must use a high-quality observation instrument to focus the observation and gather the information needed. Designing or adapting an instrument requires careful consideration of the instrument purpose, content and structure.

This annex includes a list of recommendations to support initial instrument design and adaptation, as well as the review and verification process. The content is drawn from the Classroom Observation Toolkit for Early Grade Reading Improvement. The purpose of this annex is to provide a quick reference checklist to support the instrument design process. See the EGR Program Experiences and Guidance section in the toolkit for more detailed guidance on each of the topics below.

Key recommendations on how EGR programs can design high-quality instruments and effectively apply the information gathered from these instruments are summarized below:

✓ **Identify the purpose of conducting classroom observations.** Identifying the purpose of a classroom observation exercise guides and informs decisions regarding all aspects of instrument development and use. This includes: instrument content and structure; observer profiles and the training and support they will need; the instrument medium (e.g., paper or electronic); and dissemination and use of information collected. Therefore, identifying the purpose is a critical initial step in planning for classroom observation use.

✓ **Engage diverse stakeholders and technical experts.** Collaboration and coordination of multiple stakeholders and experts throughout the entire planning and implementation process is essential to the development of a quality instrument and a successful classroom observation exercise. While the specific purpose and context will determine which individuals to involve, those usually included are: education officials at national and sub-national levels; reading and language specialists; individuals with expertise and responsibility for providing teacher professional development; research, survey design and education measurement specialists; and program monitoring and evaluation specialists.

✓ **Develop or adapt a high-quality instrument.** Instrument development and adaptation is an iterative process. To begin, select an existing instrument or develop a new one that meets criteria for the intended purpose and is appropriate for the capacity of observers. When developing or adapting an observation instrument for an EGR intervention, consider the following questions:

  * Does the instrument align with its **PURPOSE**?
  * How will the program **USE** the information?
  * Is the instrument **VALID**? How do you know?
  * Is the instrument **RELIABLE**? How do you know?
  * Does the **FOCUS** of the instrument align with what it aims to measure?
  * Is the **ITEM AND RESPONSE** format appropriate given the instrument purpose, time and expertise available for instrument development, and observer capacity?
  * Do the **INSTRUCTIONAL CATEGORIES AND BEHAVIORS** align with the purpose and the instructional strategies that the program is interested in observing?
  * Is the instrument **FEASIBLE** to implement given observer skills and the costs and time required to administer it?
  * Is the format, font, language and sentence structure of the instrument **USER FRIENDLY**?
✓ **Verify that the instrument meets the criteria above.** Once the instrument is developed or adapted, verify that it addresses each of the above criteria. If the instrument does not meet all these criteria, it may need modifications. If changes are needed, verify that these changes still meet the relevant criteria for validity and reliability; align with the instrument’s intended purpose; and are appropriate for the context, language and observer skill level.

✓ **Pilot a draft instrument.** Piloting—or trialing a draft version of an observation tool in a setting similar to the setting in which it will be used—is a critical step to verify its validity and reliability. Piloting is also an opportunity to gather feedback from observers on the instrument and to assess the quality and consistency of observer responses. Piloting should be conducted in a small, targeted sample of schools with observers who have received some training in how to use the tool. Data gathered during the pilot should be analyzed to verify whether the instrument is gathering the desired data. Piloting results should then be used to further modify the instrument and inform observer training.

✓ **Develop protocols for the classroom observation.** Instruments need to be accompanied by a protocol or guide describing how to appropriately use the instrument and conduct classroom observation visits. The protocol should include clear explanation of instrument items and how to record information. Where relevant (e.g., when observers are instructional coaches), protocols may include guidance on how to use the instrument to inform and provide feedback to teachers.

✓ **Identify, train and support classroom observers.** Those identified to serve as observers will need training on how to appropriately conduct observations and, in some cases, how to analyze and use the information collected. Training content should align with the purpose of the observation exercise and observers’ skill levels and include multiple opportunities for observers to practice using the instrument and to be assessed on their ability to administer the instrument correctly, accurately and reliably. If the observer provides feedback to teachers, training also needs to include instruction on how to use information to inform feedback. Consider staggering training to allow adequate time for skill acquisition.

✓ **Administer the classroom observation instrument.** Important issues to consider related to instrument administration include the medium of the instrument (i.e., paper or an electronic device such as a cellular phone or tablet) and the frequency and duration of classroom observation visits. Other key considerations include the following: the instrument purpose; context; observer skill level; cost of hardware, software and technical support; data management and storage; and sustainability, if relevant. While classroom observations are being administered, monitor observers to verify they are administering the instrument as intended and address any concerns accordingly.

✓ **Disseminate and use observation results.** Information collected from classroom observation instruments is only useful if it is appropriately shared and used. When identifying when, how and with whom to share findings, consider the following: purpose and audience for information sharing; strategies for communicating findings; frequency for disseminating findings; and human and other resources needed to share results.

For more detailed guidance on each of these recommendations, refer to the guidance provided in this resource, *Classroom Observation Toolkit for Early Grade Reading Improvement.*
Annex B. Identifying Teacher Professional Development Needs

DESCRIPTION OF THE TEMPLATE

Purpose and Use: The purpose of this instrument template is to support the identification of teacher professional development needs prior to developing an EGR intervention. This instrument is designed to help understand existing teaching practices and classroom conditions of teachers in a specific context to identify areas where teachers could benefit from an evidence-based early grade reading professional development intervention. Observations to assess teacher professional development needs are one element of a comprehensive needs assessment that collects data from a variety of sources. (Other data sources include: interviews and surveys on teacher attitudes, beliefs, pedagogical and content knowledge; interviews with school directors; desk reviews of teacher professional development curriculum, practices and policies; and other types of observation data). Instruments used to inform the design of an intervention might differ from instruments used in assessing teachers’ on-going training and support needs (e.g., instruments used in baseline assessments, coaching, regular program monitoring or evaluations).

Observers: Program designers, personnel or individuals hired and trained on the use and scoring of the instrument as well as how to identify specific observation items.

Item Types: This template provides examples of low-inference items, which are factual statements or questions based on behavior or classroom conditions that the observer can see. Low-inference items are easier for observers without expertise in basic pedagogy or reading instruction to accurately record a response because they do not require an evaluative judgment. The use of low-inference items maximizes the instrument’s potential for achieving higher reliability, or consistent scoring across observers on a given item. In situations where observers have more expertise, programs may choose to use high-inference items that require an evaluative judgement.

Response Format: This template uses a binary response format with a simple yes/no response option. This format allows an observer with limited expertise to check if specific behaviors have been observed or if specific classroom conditions are present. If the program is interested in assessing the frequency (how often) or duration (how long) of certain practices and behaviors, the instrument will require a different response format (e.g., objective rating scales such as yes, partially, no, or a time-sampling format). To understand the quality of instruction, the instrument would need to include high-inference items and would require an evaluative rating scale (e.g., 1-4) as a response format. These response formats, however, require a rubric with clear scoring criteria for each response option, and more extensive training on how to score reliably.

Instrument Focus: The focus of this template is to capture information on teachers’ use of general instructional strategies and literacy practices designed to improve early grade reading outcomes.

Instructional Categories and Observable Behaviors: The content of this template includes:

- Evidence-based instructional categories relevant to EGR programs (e.g., lesson structure, content and facilitation; classroom management; classroom physical environment; supportive learning environment; student engagement; checking for understanding; and feedback).
- General observable behaviors and classroom conditions relevant to skills a future EGR intervention might be targeting (e.g., systematic and explicit instruction in literacy; instructional routines, etc.).
**NEEDS ASSESSMENT: IDENTIFYING EXISTING TEACHING PRACTICES AND CONDITIONS**

This template provides a possible structure for a classroom observation instrument for identifying teacher professional development needs prior to developing an intervention. This example is tailored for observers who may not be experts in pedagogy or reading instruction. It is designed to collect information on the presence or absence of observable teacher behaviors only.

This is an illustrative template only. Programs will need to adapt the format and develop content that aligns with a specific EGR intervention and information needs. Examples are neither suggestive nor exhaustive; they are provided to illustrate instrument format and types of low-inference (e.g., yes/no) items only. The sequencing of items does not indicate the relative importance of the items or topics.

### Instructional Categories and Observable Behaviors

<table>
<thead>
<tr>
<th>Low-inference items describe a specific behavior or condition to be observed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A binary response format (e.g., yes/no) allows an observer with limited expertise to record what strategies the teacher is using and/or what classroom conditions are present.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson Structure, Content and Facilitation</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teacher has lesson notes or a teachers' guide.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Teacher connects lesson to prior learning from previous lesson (e.g., “Yesterday, we learned…”).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Teacher states the learning goal/topic of the lesson (e.g., “Today we are going to…”).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Teacher models a specific literacy skill (e.g., “I do”/modeling of reading, writing, speaking or listening).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Teacher models literacy skill while students actively participate (e.g., “We do”/guided practice of reading, writing, speaking or listening).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.
### NEEDS ASSESSMENT: IDENTIFYING EXISTING TEACHING PRACTICES AND CONDITIONS

<table>
<thead>
<tr>
<th>Instructional Categories and Observable Behaviors</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Teacher gives students an opportunity to practice literacy skill independently (e.g., “You do/ independent practice of reading, writing, speaking or listening”).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Teacher uses more than one visual aid(s) or resource to support instruction (e.g., alphabet chart, big book, bottle caps, slates, flash cards, manipulatives, local materials, etc.).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Classroom Management**

| 8. Teacher starts class on time (e.g., within X minutes of expected time on schedule). |     |    |
| 9. Teacher adheres to time allocation for specific lesson activities in accordance with teacher’s guide (if provided). |     |    |
| 10. Teacher uses at least two different types of instructional grouping (e.g., group work, pair work, individual work). |     |    |
| 11. Teacher sets behavioral expectations for students (e.g., demonstrates one of the following: states expected behavior before an activity; posts classroom rules in learning space; or establishes routines). |     |    |
| 12. Students are on task and focused on the lesson X% of the time during the observation. |     |    |

**Physical Environment**

| 13. Classroom or learning space is physically safe for ALL students (e.g., no visible risks that could cause harm). |     |    |
| 14. At least two examples of text or print are visible in the learning space (e.g., alphabet strip, pocket chart, a word wall, labels, student names, classroom rules, etc.). |     |    |
| 15. Textbooks are available for at least X% of students. |     |    |
| 16. At least one type of teaching and learning material is available for student use that is not a text book (e.g., flash cards, games, slates). |     |    |

*Low-inference items should be distinct, observable, and specific. Where needed, include a definition within text or a separate protocol with clear explanations as to what qualifies as a “yes.” On items where relevant (e.g., such as those marked with an “X” in this template), specify percentage or an amount of time based on what is expected or acceptable in the context.*

*This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.*
### Instructional Categories and Observable Behaviors

#### Classroom Culture/Supportive Learning Environment

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>17.</td>
<td>Teacher interacts with ALL students in a positive manner throughout the entire lesson.</td>
</tr>
<tr>
<td>18.</td>
<td>Teacher provides equal opportunities to ALL students to participate (e.g., girls, boys, students with different abilities and needs, students seated in different parts of the classroom).</td>
</tr>
</tbody>
</table>

#### Student Engagement

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>19.</td>
<td>Students have an opportunity to practice reading and/or writing (not copying) independently for at least X minutes [to be determined by program].</td>
</tr>
<tr>
<td>20.</td>
<td>Students have an opportunity to interact and work with their peers in at least one activity.</td>
</tr>
<tr>
<td>21.</td>
<td>Students have at least one opportunity to use learning materials.</td>
</tr>
<tr>
<td>22.</td>
<td>Students ask questions or seek assistance from teacher when needed.</td>
</tr>
</tbody>
</table>

#### Checking for Understanding

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>23.</td>
<td>Teacher checks for understanding by calling on students to answer questions or demonstrate skill.</td>
</tr>
<tr>
<td>24.</td>
<td>Teacher walks around the room and checks students work during independent, pair or group work.</td>
</tr>
<tr>
<td>25.</td>
<td>When students do not understand, teacher adjusts instruction to help them understand (e.g. rephrases questions, re-teaches concept).</td>
</tr>
</tbody>
</table>

#### Feedback

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>26.</td>
<td>Teacher praises ALL students for correct responses and behaviors. (e.g., girls, boys, students with different abilities and needs)</td>
</tr>
<tr>
<td>27.</td>
<td>When students misunderstand or perform incorrectly, teacher responds with corrective feedback in a positive manner.</td>
</tr>
</tbody>
</table>

---

This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.
Annex C. Supporting Individual Teacher Growth

DESCRIPTION OF THE TEMPLATE

**Purpose and Use:** The purpose of this instrument template is to assist coaches responsible for providing feedback and on-going support to teachers. This template is designed to help coaches collect evidence or examples of teachers’ use of instructional strategies and identify a teacher’s strengths and areas for improvement. Information collected informs feedback and reflection between the observer and teacher. After each lesson, the teacher and coach do the following: reflect on the lesson together; identify strengths and areas of improvement; establish improvement goals; and develop an improvement action plan. The information collected also informs what additional support the teacher might need. Observation data collected to support teacher growth should not be used to evaluate teacher performance or to make decisions about teacher employment status, deployment or salary.

**Observers:** Individuals trained to provide feedback and on-going instructional support to teachers (e.g., coaches, instructional specialists, pedagogical advisors, peer teachers, etc.).

**Items Types:** This template provides examples of low-inference items, which are factual statements or questions based on behavior or classroom conditions that the observer can see. Low-inference items are easier for observers without expertise in basic pedagogy or reading instruction to accurately record a response because they do not require an evaluative judgment. The use of low-inference items maximizes the instrument’s potential for achieving higher reliability, or consistent scoring across observers on a given item. In situations where observers have more expertise, programs may choose to use high-inference items that require an evaluative judgment.

**Response Format:** This template uses a binary response format with a simple yes/no response option and a space for open-ended notetaking to document evidence or examples for each behavior observed. This format allows an observer (e.g., a coach) with limited expertise to check if specific behaviors have been observed or if specific classroom conditions are present. When coaches are highly trained and have expertise in classroom instruction, instruments that are more qualitative, open-ended or evaluative of teacher performance might be more appropriate. (If evaluative rating scales are used, they require a well-developed rubric with clear criteria for how to score each item. Coaches must be trained extensively and their inter-rater reliability (IRR) assessed to ensure consistent scoring across observers and observations).

**Instrument Focus:** The focus of this template is to capture information on the following key areas:

- Teachers’ use of instructional strategies and literacy practices designed to improve early grade reading.
- Evidence or examples of behavior or instructional practice observed.
- Reflection and analysis of teacher strengths and needs.
- Action plan for improvement.

**Instructional Categories and Behaviors:** The content of an observation instrument used to support teacher growth should include:

- Evidence-based instructional categories (e.g., lesson structure, content, and facilitation; classroom management; classroom physical environment; supportive learning environment; student engagement; checking for understanding; and feedback.
- Observable behaviors and classroom conditions specific to the professional development intervention.
Note:

The observation instrument in this annex is only a template with illustrative examples of format and content. Programs will need to develop or adapt a high-quality instrument to address the specific interests of the intervention. Several factors need to be considered: observer capacity; instrument validity and reliability; instrument focus and information needs; item and response formats; instructional categories and behaviors; and feasibility and usability.

See EGR Program Experiences and Guidance in the Classroom Observation Toolkit for Early Grade Reading Improvement for specific guidance on instrument features by purpose: instrument development and adaptation; piloting; developing protocols on instrument use and scoring; identifying, training and supporting observers; administration of the instrument; and dissemination and use of results.

EGR interventions that provide on-going support to teachers through coaching require specific training for coaches beyond the use of the observation instrument. For more information on preparation and support of coaches, see Coaching in early grade reading programs: Evidence, experiences and recommendations. A Global Reading Network Resource available at www.globalreadingnetwork.net and www.education-links.org.
SUPPORTING INDIVIDUAL TEACHER GROWTH THROUGH COACHING

This template provides a recommended structure and format for a classroom observation instrument used to support teacher professional growth through coaching.

This template is informed by the following resources: Clair, N. and Hertz, A. C. (2013), Improving Schooling in Sierra Leone: Resource Manual for Learning Coaches; FHI 360's Ghana Learning’s Literacy Practice Walk Through Tool; and RTI International’s READ Liberia’s Grade 1-2 Classroom Observation Instrument.

This is NOT a complete instrument; it is an illustrative template only. Programs will need to adapt the format and develop content that aligns with the specific design of the EGR intervention and information needs. Examples are neither suggestive nor exhaustive; they are provided to illustrate instrument format and types of low-inference questions only. The sequencing of items does not indicate the relative importance of the items or topics.

| A. GENERAL INFORMATION: This section records general information needed by the program. |
|---|---|---|---|
| Name of Coach: | Date: | School Name: |
| Name of Teacher: | Class Level: | Language of Instruction: |
| Observation Start Time: | Observation End Time: | # of Students in Attendance: ___G / ___B |

<table>
<thead>
<tr>
<th>B. PRE-OBSERVATION DISCUSSION: Coach and teacher complete before the observation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>This section should provide guidance for how coaches can facilitate a pre-observation discussion with the teacher. This might be a specific script coaches can follow, talking points or a list of reflective questions coaches can ask teachers to prompt their thinking about what they want coaches to observe. This can be directly on the instrument or part of a separate document (e.g., the protocol for classroom observation visits).</td>
</tr>
</tbody>
</table>

Instructional categories and observable behaviors should reflect instructional content and behaviors specific to the EGR professional development intervention.

- **Prompts/Reflective Questions:** List specific guidance on what coaches should say, ask or do during the pre-observation discussion. Questions might refer to the previous post-observation visit and action plan for improvement and teacher progress toward the improvement goals since the last coach observation.

- **The Objective of the Lesson:** Provide space for the coach to write the objective of the lesson.

- **Observation Focus:** Provide space for the coach to write the areas for improvement/professional growth that are the focus of this observation.
### C. CLASSROOM OBSERVATION

This section is for coaches to record behaviors observed and to write specific examples observed.

<table>
<thead>
<tr>
<th>INSTRUCTIONAL CATEGORIES AND OBSERVABLE BEHAVIORS</th>
<th>YES</th>
<th>NO</th>
<th>EVIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lesson Strategies, Instructional Activities and Delivery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Teacher connects lesson to prior learning from previous lesson.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Additional items to be identified by program.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Classroom Management</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Teacher starts class on time [define, e.g., within 3 minutes of when the time period begins].</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical Classroom Environment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Items to be identified by program.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Classroom Culture/Supportive Learning Environment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Items to be identified by program.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Student Engagement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Items to be identified by program.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Checking for Understanding</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Items to be identified by program.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Items to be identified by program.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Low-inference items** describe a specific behavior to be observed. Low-inference items should be distinct, observable, and specific. Where needed, include a definition within text or a separate protocol with clear explanations as to what qualifies as a “yes.”

- The suggested **response format** for tools used to support teacher growth in EGR interventions is a binary response format with space for open-ended notetaking. This space allows the coach to document specific evidence or examples to inform the post-observation discussion with the teacher.

- Instructional categories and observable behaviors should reflect instructional content and behaviors specific to the EGR professional development intervention.
### D. COACH’S REFLECTION ON LESSON:
Prior to the post-observation visit, the coach analyzes notes from the observation and identifies teacher strengths and needs. This section includes a space for coaches to document their reflections on the lesson and identify a teacher’s strengths and areas of improvement, using specific examples to support analysis and feedback.

<table>
<thead>
<tr>
<th>STRENGTHS: Based on the evidence collected, what are the teacher’s strengths?</th>
<th>EVIDENCE: What specific examples of the teacher’s behavior from the lesson can you share with the teacher concerning the teacher’s strengths?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
<td>2.</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

**Section D** should be field tested along with the rest of the instrument to identify how long it takes to complete, the quality of the information recorded, and utility to coaches and teachers in using it.

<table>
<thead>
<tr>
<th>AREAS OF IMPROVEMENT: Based on the evidence collected, what areas can the teacher improve?</th>
<th>EVIDENCE: What specific examples of the teacher’s behavior from the lesson can you share with the teacher concerning areas in need of improvement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
<td>2.</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
E. POST-OBSERVATION DISCUSSION: Coach and teacher discuss the observation and complete this section together in the post-observation visit.

Note: This section is an example of format and content that can be included to support a post-observation discussion between a coach and teacher. It provides a space to record the outcome of the post-observation visit and for the coach and teacher to identify what support the teacher needs from the coach to help achieve improvement goals. Coaches and teachers can review the goals over time to identify what progress has been made and what areas the teacher might be struggling to improve (and therefore require more intensive support).

1. What was the teacher’s improvement goal for this lesson? Refer to Pre-Visit Discussion, Section B

2. Teacher’s reflection on the lesson: Include suggested questions or prompts for the coach to guide teacher reflection. For example: What does the teacher consider to be his or her strengths based on the lesson? How does the teacher know? Did the teacher feel students accomplished the learning objectives? How does the teacher know?

3. Coach’s recommendation on priority areas for improvements: Based on observations and reflection on the lesson, identify 1 to 2 areas you recommend the teacher focus on improving.

4. Teacher goals: List 1-2 specific goals the teacher has for improving future classroom instruction.

5. What actions would the teacher like to take to achieve this goal and by when? OR: What actions does the teacher plan to take?

6. What additional support does the teacher need? (e.g., follow-up visit, co-teaching, modeling, additional resources, etc.)

Post-observation discussion. The content of this section will vary depending on the role of coaches in the EGR intervention and the protocol established for the post-observation discussion. In this template, this section includes space for coaches who are responsible for working with teachers on goal-setting to record information about priority areas for improvement, among other areas. If coaches do not provide on-going support beyond observations and feedback, this section could be simplified based on the expectations for the post-observation discussion.

Signature of Teacher: ___________________________  Signature of Coach: ___________________________
Annex D. Monitoring Teachers’ Adherence to EGR Program Approach

DESCRIPTION OF THE TEMPLATE

Purpose and Use: The purpose of this observation instrument template is to support monitoring of EGR program implementation. This instrument is designed to help understand whether teachers are conducting lessons with fidelity, e.g., in accordance with teachers’ guides and other materials provided by the program. Information collected helps relevant stakeholders to understand if implementation is taking place as planned and if the intervention is leading to progress on predetermined indicators. The information collected can inform adaptation to the intervention as needed. This instrument does not help understand why a program is or is not implemented with fidelity.

Observers: Program personnel, monitoring and evaluation specialists, or education officials responsible for monitoring who are trained on the use and scoring of the instrument as well as how to identify specific observation items.

Item types: This template provides examples of low-inference items, which are factual statements or questions based on behavior or classroom conditions that the observer can see. Low-inference items are easier for observers without expertise in basic pedagogy or reading instruction to accurately record a response because they do not require a valuative judgement. The use of low-inference statements maximizes the instrument’s potential for achieving higher reliability, or consistent scoring across observers on a given item. In situations where observers have more expertise, programs may choose to use high-inference items that require an evaluative judgement.

Response Format: This template uses a binary response format with a simple yes/no response option. This format allows an observer with limited expertise to check if specific behaviors have been observed or if specific classroom conditions are present. If the program is interested in assessing the frequency (how often) or duration (how long) of certain practices and behaviors, the instrument will require a different response format (e.g., objective rating scales such as yes, partially, no, or time-sampling formats). To understand the quality of instruction, the instrument would need to include high-inference items and would require an evaluative rating scale (e.g., 1-4) as a response format. These response formats, however, require a rubric with clear scoring criteria for each response option and more extensive training on how to score reliably.

Instrument Focus: The focus of this template is to capture information on the following key areas:

- Teachers’ adherence to strategies or behaviors the intervention expects teachers to be doing for a given lesson (e.g., implementing the “I do, we do, you do” approach to instruction).
- Availability and use of materials provided by the intervention (e.g., teacher’s guide).
- Other information needed for monitoring purposes (e.g., attendance, GPS).

Instructional Categories and Observable Behaviors: The content of this template includes:

- Evidence-based instructional categories (e.g., lesson structure, content and facilitation; classroom management; classroom physical environment; supportive learning environment; student engagement; checking for understanding; and feedback).

Note:

Depending on a program’s monitoring needs, the organization of instruments may differ considerably. Some may organize content by steps in a lesson rather than by instructional categories, as illustrated in this template. Some might tailor items to focus explicitly on adherence to specific lessons as prescribed in a teacher’s guide. Some might also include sections with items for the observer to circle specific activities observed, or to record information such as the number of students who have books. Whatever approach is decided, the content needs to align with the specific monitoring purpose and what it measures.
The observation instrument provided in this annex is only a template with illustrative examples of format and content. Programs will need to develop or adapt a high-quality instrument to address the specific interests of the intervention. Several factors need to be considered: observer capacity; instrument validity and reliability; instrument focus and information needs; item and response formats; instructional categories and behaviors; and feasibility and usability.

See EGR Program Experiences and Guidance in the Classroom Observation Toolkit for Early Grade Reading Improvement for specific guidance on instrument features by purpose: instrument development and adaptation; piloting; developing protocols on instrument use and scoring; identifying, training and supporting observers; administration of the instrument; and dissemination and use of results.

MONITORING FIDELITY OF IMPLEMENTATION: TEACHERS’ ADHERENCE TO EGR PROGRAM APPROACH

This template is an example of a structure for a classroom observation instrument for monitoring teachers’ adherence to guided lesson plans (or teacher’s guide) and use of teaching and learning materials in a specific EGR intervention. This example only monitors adherence (whether teachers are using strategies); it does not monitor other aspects of implementation fidelity including the frequency, duration or quality of teachers’ use of strategies.

This template and content are informed by several instruments from EGR programs used for monitoring adherence, including: READ Liberia (RTI International); Nigeria Reading and Access Research Activity (RTI International); Nigeria Northern Education Initiative Plus (Creative Associates); Madagascar Mahay Mamoky Teny (FHI 360); and Ghana: Learning (FHI 360).

This is an illustrative template only. Programs will need to adapt the format and develop content that aligns with the specific design of the EGR intervention and information needs. Examples are neither suggestive nor exhaustive; they are provided to illustrate instrument format and types of low-inference questions only. The sequencing of items does not indicate the relative importance of the items or topics.

### A. GENERAL INFORMATION

Date: |___|___|___|___20|___|
Observer Name: ____________________________
School Name: ____________________________ GPS Location: ____________
Teacher’s Name: (if required, or unique identification code) ______________
Gender of Teacher: ______
Grade: ______ Lesson Week: ____________ Lesson Day: _____________
Language(s) of lesson: __________ Language(s) Used for Instruction: __________
Number of Students Enrolled in Class: Girls: |___|___| Boys: |___|___|
Number of Students Attending Class Today: Girls: |___|___| Boys: |___|___|
If fewer than 10 pupils are in class today, write a 0 before the number. Example: 06
Expected Duration of Lesson: |____|____| minutes
Start Time of Lesson: |____|____: |____|___ am/pm End Time of Lesson: |____|____: |____|___ am/pm
Total Time Spent Delivering Lesson: |____|____| minutes

General information: Observation instruments for program monitoring may require more specific and detailed information than other types of instruments. This will depend on the information the program needs for monitoring implementation.

This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.
# B. CLASSROOM OBSERVATION

## Instructional Categories and Observable Behaviors

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
</table>

### Lesson Structure, Content, and Facilitation

1. Teacher shows evidence of preparedness [list program-specific items e.g., materials are prepared, teacher has song for today's lesson written on the board].

2. Teacher connects lesson to prior learning from previous lesson.

3. Teacher states the learning goal/topic of the lesson (e.g., “Today we are going to…”).

4. “I do” - Teacher follows teacher’s guide as indicated and models specific literacy skill or activity (e.g., modeling of reading, writing, speaking or listening skill).

5. “We do” - Teacher follows teacher’s guide as indicated and models literacy skill or activity with ALL students actively participating together with the teacher (e.g., guided practice of reading, writing, speaking or listening skill).

6. “You do” - Teacher follows teacher’s guide as indicated and provides students an opportunity to practice literacy skill independently (e.g., independent practice of reading, writing, speaking, or listening).

### Classroom Management

7. Teacher starts class on time (e.g., within X minutes of expected time on schedule).

8. Teacher adheres to time allocation for specific lesson activities in accordance with teacher’s guide (if provided).

### Physical Classroom Environment

9. Classroom or learning space is physically safe for ALL students (e.g., no visible risks that could cause harm).

10. Every student has learning materials specified and supplied by the program [list materials].

11. At least X% of student books are in usable condition (e.g., no missing or torn pages).

12. All other teaching and learning materials provided by the program are available for student use [list program-specific items such as big books, posters, flash cards, games].

---

This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.
### B. CLASSROOM OBSERVATION

#### Instructional Categories and Observable Behaviors

<table>
<thead>
<tr>
<th>Classroom Culture/Supportive Learning Environment</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Teacher provides equal opportunities to ALL students to participate (e.g., girls, boys, students with different abilities and needs, students seated in different parts of the classroom).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Student Engagement

| 14. Students practice reading and/or writing (not copying) independently for at least X minutes [to be determined by the program]. |     |    |
| 15. Students use the learning materials provided by the program in at least one activity. |     |    |
| 16. Students ask questions or seek assistance from teacher when needed. |     |    |
| 17. Students point to the text as they read in their student workbooks (to demonstrate tracking). |     |    |
| 18. All students complete the day’s exercises as specified in the teacher’s guides. |     |    |

#### Checking for Understanding

| 19. Teacher checks student understanding throughout the lesson as indicated in the teacher’s guide [list program-specific behaviors expected for this item, such as: calls on students or walks around the room to check students’ work]. |     |    |
| 20. Teacher adjusts instruction to meet the needs of ALL students when they do not understand [provide definition of what this should look like based on intervention, e.g., rephrases questions, re-teaches skill if needed]. |     |    |

#### Feedback

| 21. Teacher praises ALL students for correct responses and behaviors (e.g., girls, boys, students with different abilities and needs). |     |    |
| 22. When students perform incorrectly, teacher responds with positive corrective feedback. |     |    |

---

This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.
Annex E. Measuring Changes in Teachers’ Use of Instructional Time

DESCRIPTION OF THE TEMPLATE

Purpose and Use: The purpose of this observation instrument template is to help an intervention understand teachers’ use of instructional time on EGR program-specific content and strategies. Specifically, this instrument collects information on the use of instructional time on reading skills, teacher-related tasks and student engagement. Governments or programs can use this instrument for different observation purposes. However, in the context of EGR interventions, this instrument is commonly used in evaluations. Information collected helps the program, government and donor to understand the extent to which changes in teachers’ use of instructional time occur because of the intervention. In some cases, programs may choose to analyze observation data with EGRA results to identify if a correlation exists between teachers’ use of instructional time and improved student learning outcomes.

Observers: Evaluators, enumerators, or other individuals who are hired, trained, and formally assessed on the accurate and reliable use of the instrument.

Item types: This template provides examples of low-inference items, which are factual statements or questions based on behavior or classroom conditions that the observer can see. Low-inference items are used because they are easier for observers without expertise in basic pedagogy or reading instruction to accurately record a response. The use of low-inference items maximizes the instrument’s potential for achieving higher reliability, or consistent scoring across observers on a given item. In situations where observers have more expertise, programs may choose to use high-inference items that require an evaluative judgement.

Response Format: This template uses a simplified time-interval sampling format—often referred to as a “timed instrument” or “classroom snapshot.” The format on this template allows observers to collect information on estimated use of instructional time on reading content, teacher-related tasks, and student engagement. If programs are interested in collecting other types of information related to time (e.g., use of time on specific teaching and learning materials, time on teacher-student interactions, or ability of teachers to keep students on task and engaged), the instrument will require a different item and response format. See Stallings classroom observation system, or classroom snapshot, for an example of an alternative approach and more information.

Instrument Focus: The focus of this template is to capture information on the following areas:

- Teachers’ use of instructional time on content (e.g., foundational skills in reading).
- Teachers’ use of instructional time on teacher-related tasks over the course of a lesson.
- Student engagement in opportunities to learn.

Instructional Categories and Behaviors: The content of observation instruments used for measuring how instructional time is used should align with evidence-based instructional content (e.g., foundational skills in reading) and behaviors one would expect to see in an early grade reading lesson.

Note:
The observation instrument in this annex is only a template with illustrative examples of format and content. This instrument template has not been piloted or tested for validity and reliability for use across programs. Programs will need to develop or adapt a high-quality instrument to address the specific interests of the intervention. Several factors need to be considered: observer capacity; instrument validity and reliability; instrument focus and information needs; item and response formats; instructional categories and behaviors; and feasibility and usability.
If programs are interested in a reliable and validated instrument for use in impact evaluations to collect comparable data on use of instructional time or other time-related measures, see Stallings classroom snapshot for an example that has been used in EGR evaluations.

See EGR Program Experiences and Guidance in the Classroom Observation Toolkit for Early Grade Reading Improvement or specific guidance on instrument features by purpose: instrument development and adaptation; piloting; developing protocols on instrument use and scoring; identifying, training and supporting observers; administration of the instrument; and results dissemination. This annex provides guidance for developing one type of instrument that could be used as part of an evaluation.

**MEASURING USE OF INSTRUCTIONAL TIME IN EGR PROGRAMS**

This template is an example of a structure and format for a classroom observation instrument for measuring teachers’ use of instructional time during a lesson. This is NOT a complete instrument; it is an example of a time-sampling format that EGR programs may consider when developing instruments to collect data on use of instructional time.

This template is informed by the following resources: Nigeria Reading and Access Research Activity (RARA) Timed Classroom Observation Instrument (RTI International, 2016); Zambia Time to Learn Classroom Observation Protocol (Encompass, 2017); the TEACH Classroom Observation Tool (World Bank, 2019); and Stalling’s Classroom Snapshot Observation System (World Bank, 2015).

This is an illustrative template only. Programs will need to adapt the format and develop content that aligns with the specific design of the EGR intervention and information needs. Examples are neither suggestive nor exhaustive; they are provided to illustrate instrument format and types of low-inference questions only. The sequencing of items does not indicate the relative importance of the items or topics.

This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.
INSTRUCTIONS FOR OBSERVERS

Provide instructions here for what observers need to do prior to the lesson, during the lesson, and after the lesson.

ORGANIZATION OF THE INSTRUMENT

The observation template is subdivided into three parts:

Part I: Background Information: This section is for recording general information about the school, class and lesson. Fill in this information prior to the observation.

Part II: Classroom Observation: This section is for observers to record behaviors observed in the observation. There are four sections:

- **Section A** records the time allotted for instruction and the actual instructional time.
- **Section B** relates to subject-specific content. Six specific categories relate to foundational literacy skills (print concepts, phonemic awareness and phonics, writing, vocabulary, comprehension, and oral language), and an additional row is included for “Other.” Each category is subdivided by activities that the teacher or students are doing.
- **Section C** relates to the teacher’s actions during lesson delivery.
- **Section D** relates to student engagement during the lesson.

Each section is organized by columns into 3-minute intervals. Each column represents 3 minutes of lesson time, beginning with the first column on the far left (first three minutes after the teacher begins teaching) to the last column on the far right (the last 42 to 45 minutes). There are 15 columns for a cumulative total of 45 minutes.

Part III. Evaluation Criteria: This section includes specific criteria for each of the items being observed in Sections B, C and D.

HOW TO USE THE INSTRUMENT

The observer will observe a 45-minute lesson and record what is occurring during the lesson in 3-minute intervals. Once the teacher begins teaching every 3 minutes, the observer should find the teacher and then scan the classroom in a clockwise direction and tick the appropriate column and row of the action or activity that is observed. Once 3 minutes have elapsed, the observer will move to the next column to the right. If the teacher is not providing instruction on any of the skills listed, the observer will tick “other.”

Instruments should be accompanied by a protocol or guide explaining how to administer the instrument. For this template, Instructions for Observers provides space to include specific instructions for the observers describing what they need to do prior to the lesson, during the lesson, and after the lesson. This can include procedures for informed consent and guidance on appropriate observer etiquette.

The section on Organization of the Instrument describes the different sections of the instrument. The text for this section is illustrative only and should be revised to align with the instrument developed for the evaluation. For example, some may want to include a section on teachers’ use of instructional time using specific materials.

The section on How to use the instrument provides specific directions on how to use the instrument. The content in this section is illustrative only and should be revised to align with the instrument developed for the evaluation. For example, 45 minutes is an example time only, and should be adapted for the context. Experience has found that some teachers will teach beyond the prescribed lesson time. Thus, instruments should be designed to capture a lesson that lasts longer than may be prescribed in the lesson plan. Also, for instruments that are completed electronically, the interface and instructions to observers will be different.
## I. BACKGROUND INFORMATION
*Items to be identified by the program.*

## II. CLASSROOM OBSERVATION

### SECTION A: Duration of Lesson Observation

| Expected Start Time: | ———— | ———— |
| Actual Start Time: | HH | MM |

| Expected End Time: | ———— | ———— |
| Actual End Time: | HH | MM |

### SECTION B: Use of Instructional Time on Specific Reading Skills (What is the content focus of the instruction?)

Select the reading skills the teacher is teaching for each 3-minute interval. **Note:** You may select MORE THAN ONE item for each 3-minute interval.

<table>
<thead>
<tr>
<th>Reading Skills</th>
<th>Specific Content</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
<th>15</th>
<th>18</th>
<th>21</th>
<th>24</th>
<th>27</th>
<th>30</th>
<th>33</th>
<th>36</th>
<th>39</th>
<th>42</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Print Concepts</td>
<td>1.1. Directionality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.2. Items to be identified specific to the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Phonemic Awareness and Phonics</td>
<td>2.1. Letter sounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.2. Items to be identified specific to the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Passage/Story Reading</td>
<td>3.1 Read-aloud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.2. Items to be identified specific to the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Writing (Student writing)</td>
<td>4.1. Drawing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.2. Items to be identified specific to the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Vocabulary and Reading Comprehension</td>
<td>5.1 Vocabulary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.2. Items to be identified specific to the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Oral Language</td>
<td>6.1. Listening Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.2 Items to be identified specific to the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Other</td>
<td>7.1 Items to be identified specific to the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**The time-interval sampling format of Sections B and D allows observers to record what instructional content and strategies are present every three minutes.**

**The content of Section B should align with the specific foundational literacy skills reinforced in the EGR teacher professional development intervention. The content listed in this template (e.g., directionality, letter sounds) is only illustrative and needs to be modified and expanded to reflect the skills specific to the intervention being evaluated.**

---

This template is not intended for direct use in a program. Refer to the Classroom Observation Toolkit for Early Grade Reading Improvement for guidance when developing instruments.

Section I should include any general or background information that is needed for the evaluation concerning the school, classroom, lesson, or teacher (e.g., date of observation, school name, location, date, class, grade, and language(s) used for instruction).
**SECTION C: Use of Instructional Time in Lesson Delivery (How is the teacher delivering the lesson?)** Select the item that best describes what the teacher is doing during each 3-minute interval. **Note:** You may select ONLY ONE item per 3-minute interval.

- 8. Teacher provides direct instruction to the whole class
- 9. Teacher assigns activities for students to practice in pairs or groups
- 10. Teacher assigns activities for students to practice independently.
- 11. Teacher walks around room to check student work.
- 12. Teacher is managing student behavior but not teaching.
- 13. Teacher is in the classroom but not interacting with students.
- 14. Teacher is not in the room.
- 15. Items to be identified specific to the program.

**Low-inference statements** describe a specific behavior to be observed. Low-inference items should be distinct, observable, and specific.

**Section C** captures teacher-related tasks during lesson delivery. The example content is only illustrative and should be modified to reflect the instructional strategies specific to the EGR intervention being evaluated. For example, some instructional time instruments may focus on the use of instructional routines. Others may choose to capture whether lesson delivery involves whole class instruction, groupwork, pair work or independent work.

**SECTION D: Student Engagement during the lesson (To what extent are students engaged in learning during the lesson?)** Select the item that best describes how students are engaged during each 3-minute interval. **Note:** You may select ONLY ONE item per 3-minute interval.

- 16. Students listen to the teacher.
- 17. Students work in small groups or pairs.
- 18. Students practice reading or writing (not copying) independently.
- 20. Items to be identified specific to the program.

**Section D** captures the extent to which students are on-task and engaged in learning throughout the lesson.

Both **Section C** and **Section D** together help understand the extent to which students are provided with opportunities to learn and the extent to which students engage with those opportunities.
### III. EVALUATION CRITERIA OF OBSERVATION ITEMS

<table>
<thead>
<tr>
<th>Specific Reading Skill</th>
<th>Description of Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.1 Directionality</strong></td>
<td>Teacher tracks text with a finger or pointer when reading out loud to class.</td>
</tr>
<tr>
<td>Items to be identified by the program.</td>
<td>To be defined</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson Delivery</th>
<th>Description of Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>8. Teacher provides direct instruction to the whole class.</strong></td>
<td>Teacher presents or demonstrates a skill or concept to whole class.</td>
</tr>
<tr>
<td>Items to be identified by the program.</td>
<td>To be defined</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Engagement</th>
<th>Description of Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items to be identified by the program.</td>
<td>To be defined</td>
</tr>
</tbody>
</table>

The instrument should include clear, specific descriptions of each item to be observed. These descriptions will support observers who may not be familiar with instructional strategies, as well as reliability among observers. The example in Part III is illustrative of a potential template only.
Annex F. Simple Observation Checklist

DESCRIPTION OF THE INSTRUMENT

Purpose and Use: The purpose of this observation instrument template is to help personnel from USAID, government, or partner organizations to gain a general understanding of whether a specific primary grade reading intervention is being implemented as planned. It is designed specifically to help understand if classrooms visited are implementing the program as intended and using specific teaching and learning materials. Information collected can be used to stimulate and inform discussions among stakeholders about program implementation. This instrument should not be used for coaching, monitoring program implementation against specific indicators, or evaluating teacher performance.

Observers: Donor, program or government personnel who may conduct periodic school visits to monitor EGR programs but are not experts in reading instruction or teacher professional development may not be familiar with the context or may not speak the language used in the classroom. For example, this may be a senior-level USAID or government staff person who occasionally visits a classroom where an early grade reading program is being implemented.

Item types: This template includes low-inference items, which are factual statements or questions based on behavior or classroom conditions that the observer can see. The number of items provided in the template are illustrative.

Response Format: This template uses a binary response format with a simple yes/no response option. Binary response formats are most appropriate for low-inference instruments and help an observer record whether a behavior or classroom condition is observed at least once in the lesson.

Instrument Focus

- Teachers’ use of specific instructional strategies and teaching and learning materials in a particular EGR improvement intervention.

Instructional Categories and Observable Behaviors: The content of this instrument includes:

- Evidence-based instructional categories common in EGR programs (e.g., lesson structure, content and facilitation; classroom management; classroom physical environment; supportive learning environment; student engagement; checking for understanding; and feedback).
- Observable behaviors and classroom conditions common in EGR programs that provide teaching and learning materials as part of the intervention.

Note:
The instrument provided is a template with examples of format and content that is ready for use for the purposes described. However, someone familiar with the intervention may choose to adjust the instrument to be more specific to the behaviors and conditions expected for a given program approach and lesson. For example, an intervention aiming to improve social and emotional learning (SEL) would want to add behaviors or conditions specific to SEL.

See Observation Instrument Features and Content in the Classroom Observation Toolkit for Early Grade Reading Improvement for more detailed information.
INSTRUCTIONS FOR OBSERVERS: HOW TO USE THE INSTRUMENT

Step 1: Before observing, review the checklist of instructional categories, observable behaviors and classroom conditions. Discuss with program personnel to verify you understand each item and how to record information about it.

Step 2: Obtain a copy of the lesson plan or teacher’s guide provided by the program.

Step 3: Closely watch what is taking place in the classroom. After you feel comfortable, start recording information about the items in the checklist. Check “yes” if the behavior is observed at least once in the observation; check “no” if the behavior is not observed.

Step 4: After observing the lesson and interactions between teachers and children, look around the classroom and record information about items in the instrument related to the physical environment.

### SIMPLE OBSERVATION CHECKLIST OF EGR LESSON

<table>
<thead>
<tr>
<th>Instructional Categories and Observable Behaviors</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lesson Structure, Content, and Facilitation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Teacher is fully prepared for the day’s lesson, as indicated in the teacher’s guide (e.g., ALL materials are prepared; teacher has song for today’s lesson written on the board).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Teacher follows teacher’s guide for modeling specific literacy skill or activity (e.g., “I do”/modeling of reading, writing, speaking or listening).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Teacher follows teacher’s guide for modeling literacy skill or activity while students actively participate with the teacher (e.g., “We do”/guided practice of reading, writing, speaking, or listening).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Teacher follows teacher’s guide for giving students an opportunity to practice literacy skill or activity independently (e.g., “You do”/independent practice of reading, writing, speaking, or listening).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Teacher uses instructional strategies and literacy practices as described in the teacher’s guide.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Teacher uses program resources as described in the lesson to model reading skills and support the lesson (e.g., teacher’s guide, student books, and other teaching and learning materials, etc.).</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Classroom Management</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Teacher starts class on time (e.g., within three minutes of expected time on schedule).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Teacher adheres to the time allocation for specific lesson activities in accordance with teacher’s guide (if provided).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Teacher uses at least two types of instructional grouping at least once during the class (e.g., whole group, small group, pair work, individual work).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Students are on task and focused on the lesson for at least 90% of the time during the observation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical Classroom Environment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Classroom or learning space is physically safe for ALL students (e.g., no visible risks that could cause physical harm).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. At least one example of text or print is visible in the learning space (e.g., alphabet strip, pocket chart, a word wall poster; labels, student names, classroom rules, student work, etc.).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# SIMPLIFIED OBSERVATION CHECKLIST OF EGR LESSON

## Instructional Categories and Observable Behaviors

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.</td>
<td>A student book provided by the intervention [e.g., primer; insert specific name for program] appears to be available for ALL students.</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>At least one type of teaching and learning material provided by the intervention is available for student use that is not a student book primer (e.g., supplementary reading books, posters, flash cards, etc.).</td>
<td></td>
</tr>
</tbody>
</table>

## Classroom Culture/Supportive Learning Environment

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15.</td>
<td>Teacher provides equal opportunities to ALL students to participate (e.g., girls, boys, students with different abilities and needs, students seated in different parts of the classroom).</td>
</tr>
<tr>
<td>16.</td>
<td>Teacher interacts with ALL students in a positive and respectful manner throughout the entire lesson.</td>
</tr>
</tbody>
</table>

## Student Engagement

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>17.</td>
<td>Students practice reading and/or writing (not copying) independently in at least one activity.</td>
</tr>
<tr>
<td>18.</td>
<td>Students interact and work with peers during at least one activity.</td>
</tr>
<tr>
<td>19.</td>
<td>Students use the learning materials provided by the program in at least one activity.</td>
</tr>
<tr>
<td>20.</td>
<td>Students seek assistance from teacher when needed (e.g., ask questions).</td>
</tr>
</tbody>
</table>

## Checking for Understanding

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>21.</td>
<td>Teacher checks student understanding throughout the lesson as indicated in the teacher’s guide (e.g., calls on students, walks around the room checking students’ work).</td>
</tr>
</tbody>
</table>

## Feedback

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>22.</td>
<td>Teacher praises ALL students for correct responses and behaviors.</td>
</tr>
<tr>
<td>23.</td>
<td>When students perform incorrectly, teacher corrects incorrect responses AND gives student(s) an opportunity to try again.</td>
</tr>
</tbody>
</table>